CONTEMPORARY RESEARCH ON TEACHER EFFECTIVENESS

BIDDLE

ELLENA

3149

Contemporary
Research
on Teacher
Effectiveness

Contributors

BRUCE J. BIDDLE
HAZEL DAVIS
WILLIAM J. ELLENA
NED A. FLANDERS
PAUL V. GUMP

MILTON MEUX
WARREN A. PETERSON
HOWARD A. ROSENCRANZ
DAVID G. RYANS
B. OTHANEL SMITH

R. L. TURNER

Contemporary Research on Teacher Effectiveness

Bruce J. Biddle
William J. Ellena
Editors

Sponsored by American Association of School Administrators,
National School Boards Association, The Department of
Classroom Teachers of the National Education Association

Holt, Rinehart and Winston
New York · Chicago · San Francisco · Toronto · London

S.C.E.R.T., West Bengal
Library
Calcutta
25/3, B. C. Road

~~EVALUATION UNIT LIBRARY
DAVID HARE TRAINING
COLLEGE
CALCUTTA-19
Acc. No. 0039
Date 30 MAR 1965
Call No. 370.78 B554~~
65/788

S.C.E.R.T., West Bengal
Date 12-3-85
Acc. No. 3149

370.78
BID

The three associations are indeed pleased to sponsor this document. It should be pointed out, however, that the interpretations and conclusions presented herein are those of the authors and editors and do not necessarily reflect the viewpoint of the American Association of School Administrators, the Department of Classroom Teachers of the NEA, or the National School Boards Association.

Copyright © 1964 by Holt, Rinehart and Winston, Inc.
All Rights Reserved

Library of Congress Catalog Card Number: 64–18454
20944–0114

Printed in the United States of America

PREFACE

Probably no aspect of education has been discussed with greater frequency, with as much deep concern, or by more educators and citizens than has that of teacher effectiveness—how to define it, how to identify it, how to measure it, how to evaluate it, and how to detect and remove obstacles to its achievement. Separate facets of this problem have been studied, too, by state and local school systems, by individuals, and by teams of educational researchers at universities. But findings about the competence of teachers are inconclusive and piecemeal; and little is presently known for certain about teacher excellence.

The problem is not an idle one. The domestic scene and world outlook both clearly demonstrate the urgent need for more and better education for all men. Of all societies, the free one depends most heavily on quality education for the fulfillment of its destiny. Thus, the teacher and the quality of his teaching are of paramount importance.

One finds general agreement that the goal should be a highly competent teacher in every classroom and correspondingly competent administrators and consulting specialists in all leadership positions in the school system. One finds considerably less agreement, however, on the meaning and evaluation of competence.

In recent years a major concern of school administrators, school board members, parents, and teachers has focused on ways and means of establishing merit salary schedules or methods to measure teacher performance. Unfortunately, discussions of teacher effectiveness are often fraught with more emotion than good sense. It is not unusual to hear such statements as, "It can be done if you have the courage" or "It is all right in theory, but not in practice" or "Such action will destroy the morale of teachers and wreck the profession." Well-known authors and speakers often appear as ardent champions of teacher evaluation or implacable foes of the menace being proposed; and many schemes for teacher evaluation have been proposed, adopted, defamed, defended, or dropped by concerned educators and school systems.

Nor have researchers been inactive. Recent summaries have revealed that literally thousands of studies have been conducted on teacher excellence since the beginning of the twentieth century. Investigators have looked at teacher training, traits, behaviors, attitudes, values, abilities, sex, weight, voice quality, and many other characteristics. Teacher effects have been judged by investigators themselves, by pupils, by administrators and parents, by master teachers, by practice teachers, and by teachers themselves. The apparent results of teaching have been studied, including pupil learning, adjustment, classroom performance, sociometric status, attitudes, liking for school, and later achievement. And yet, with all this research activity, results have been modest and often contradictory. Few, if any, facts are now deemed established about teacher effectiveness, and many former "findings" have been repudiated.

It is not an exaggeration to say that we do not today know how to select, train for, encourage, or evaluate teacher effectiveness. And many educational researchers have abandoned the field of competence research as a simple-minded approach to a vastly more complex topic: the study of classroom interaction.

Such dismal results provide little comfort for the school administrator, who is confronted with everyday, real problems in the field of teacher excellence. Practical decisions have to be made, and these decisions are dependent upon ideas about quality of teaching. For example, applicants for teaching positions must be screened, prepared, and judged. Teachers must be employed, assigned, transferred, and occasionally released. Problems such as these do not await definitive answers.

On the one hand, school officials face the practical necessity of making evaluative judgments about the work of teachers. On the other, we know that teaching is a complex phenomenon about which little is known outside the realm of personal opinion. Few, if any, problems are more crucial in education; and, as elusive as teacher excellence may be, research toward its understanding must continue.

It was with this orientation that three national associations—the

Department of Classroom Teachers of the National Education Association, the National School Boards Association, and the American Association of School Administrators—convened a seminar of educational researchers whose work, it was felt, might provide new insights on the teacher effectiveness problem. Participants in the seminar were selected because they had directed or participated in research programs representing new approaches to classroom instruction or the structure of education. Each was asked to prepare, in advance, the draft of a paper which was to be presented at the seminar. Papers were to discuss the author's own research rather than review the work of others.

The actual seminar was convened in Washington, D.C., during the summer of 1961. Papers were read, criticized, revised. Participants gained insights both about their own work and that of others. Following the conference, final papers were revised (between January and August 1962) and then edited and assembled into the present volume, together with an introductory chapter which presents a model for research on the effectiveness problem.

This book, then, concerns research approaches to the problem of teacher effectiveness. It is directed to two audiences: concerned educators and lay persons who want to do something about our understanding of teacher competence, and educational researchers who would explore new approaches to this age-old problem.

It is our hope that educators will find here an understanding of the depth and complexity of the teacher effectiveness problem; that they will be warned against facile and shortsighted solutions to the problem of teacher evaluation; and that we can enlist their leadership, support, and participation in the fascinations of research on classroom instruction.

For the educational researcher this book is intended as a departure from traditional models of educational theory and research. By examining a series of new programs in educational research, we hope to generate insights about the process of instruction and the physical and social contexts in which education is performed. Moreover, we hope to stimulate researchers into long-term commitments to research programs bearing on the competence problem.

BJB
WJE

May 1964

THE CONTRIBUTORS

Bruce J. Biddle was born in 1928, took his A.B. degree in mathematics at Antioch College, and completed his Ph.D. degree in social psychology at the University of Michigan. He is presently associate professor of psychology and sociology at the University of Missouri, where he directs the Social Psychology Laboratory. His major fields of interest are mathematical approaches to social theory, the social psychology of education, and role theory. He has conducted extensive research in the latter field, is author of numrous articles on educational roles, and will be coeditor of a forthcoming reader in role theory.

Hazel Davis is associate director of the Research Division of the National Education Association. After completing her B.S. degree at Teachers College, Columbia University, her M.A. degree at the University of Chicago, and her Ph.D. degree at Teachers College, Columbia University, she joined the staff of the NEA, where she has been employed for the past twenty-seven years. She has edited numerous yearbooks and has contributed frequently to NEA publications as well as to many other national journals and periodicals.

William J. Ellena is associate secretary of the American Association of School Administrators, a position he has held for the past five years. Born in Minnesota in 1929, he took his B.S. and M.A. degrees at the University of Minnesota and completed his doctorate at the University of Maryland. He has been a teacher, professor, administrator, and researcher, and has been a frequent contributor to national magazines as well as a contributing author to several books on rural education and school administration.

Ned A. Flanders was born in 1918 and took undergraduate degrees at the University of California and Oregon State University. He also received an M.A. degree from the latter institution before moving to Chicago for his Ph.D. degree in education. He has had extensive teaching experience in both public schools and at the University of Minnesota, where he served until 1962. Presently he is professor of education and consultant on research at the University of Michigan. His special interests focus on techniques for observation of classroom behavior, group dynamics, and social-influence theory. He is author of many articles on classroom interaction and the use of interaction tools for retraining teachers.

Paul V. Gump was born in 1920 and took the majority of his education at Ohio State University, where he received a bachelor's degree in education and a master's degree in psychology. His Ph.D. degree was granted from the University of Colorado, in clinical psychology, and he has had extensive experience as a school psychologist and counselor. His major interests are in children's games, classroom interaction, and psychological ecology. He has written in each of these fields. He was associate professor of education at Wayne State University until 1962; he recently accepted a similar position in psychology at the University of Kansas.

Milton Meux was born in 1930 and received his B.S. degree from Kalamazoo College in Michigan and his Ph.D. degree in the field of psychology at the University of Illinois. His special interest is in the area of psychometrics, and he has written a number of articles in this area. Currently he is assistant research professor in the Bureau of Educational Research at the University of Illinois.

Warren A. Peterson is director, Department of Social Problems and Education, Community Studies, Inc., in Kansas City, Missouri. He was born in 1922, took his B.S. degree at Western Michigan College and his M.A. degree and Ph.D. degree in sociology at the University of Chicago. He has conducted extensive research in the field of psychology-sociology and is a frequent contributor to scholarly journals.

Howard A. Rosencranz was born in 1920 and received his A.B. degree from Winona State College in social science. His M.A. degree (psychology) and Ph.D. degree (sociology) were both from Michigan State, where he taught until 1960. He is presently research associate and associate professor of sociology and psychology at the University of Missouri, where he teaches social psychology and the sociology of education. He has done research at both Michigan State University and the University of Missouri on teacher role and has published in this field and has contributed to texts on introductory sociology.

David G. Ryans is presently head of the Center for Research in Systems Development and head of education research and development at the Systems Development Corporation in Santa Monica, California. He was born in 1909 and received his undergraduate education at DePauw University. His M.A. and Ph.D. degrees were awarded at the University of Minnesota, and for several years he taught psychology. He has also been director of national teacher examinations for the American Council on Education and professor of educational psychology at both the University of California in Los Angeles and the University of Texas, Austin. His major field of interest is in behavioral research and theory, in which he has published extensively, including his most famous book *Characteristics of Teachers*. He is the immediate past president of the American Educational Research Association.

B. Othanel Smith took his B.S. degree from the University of Florida and his M.A. and Ph.D. degrees at Columbia University. After serving as both a classroom teacher and a supervising principal, he joined the faculty at the University of Florida. Since 1945 he has served as professor of education at the University of Illinois, where he has conducted research on curriculum, the logic of educational measurements, social foundations of education, and language and concepts in education. He has written a wide variety of articles and books and is the past president of the National Society of College Teachers of Education.

R. L. Turner was born in 1929, took his B.S. and M.A. degrees from Northwestern University, and completed his Ph.D. degree at Indiana University, where he is assistant professor in the Institute of Educational Research. He has been a supervising teacher and taught at one time at Antioch College and at Wisconsin State College in Whitewater. His major interests are in research in teacher behavior, problem solving, and the transfer of training; he has published a variety of articles on these problems.

CONTENTS

Preface v

The Contributors ix

1 The Integration of Teacher Effectiveness Research BRUCE J. BIDDLE 1

2 Evolution of Current Practices in Evaluating Teacher Competence HAZEL DAVIS 41

3 Research on Teacher Behavior in the Context of the Teacher Characteristics Study DAVID G. RYANS 67

4 Teaching as Problem-Solving Behavior: A Strategy R. L. TURNER 102

5 Logical Dimensions of Teaching Behavior MILTON MEUX AND B. OTHANEL SMITH 127

6 Environmental Guidance of the Classroom Behavioral System PAUL V. GUMP 165

7 Some Relationships among Teacher Influence, Pupil Attitudes, and Achievement NED A. FLANDERS 196

8 The Role Approach to Teacher Competence HOWARD A. ROSENCRANZ AND BRUCE J. BIDDLE 232

9 Age, Teacher's Role, and the Institutional Setting WARREN A. PETERSON 264

Bibliography 316

Index 345

I

BRUCE J. BIDDLE

The Integration of Teacher Effectiveness Research

This book concerns research on the effectiveness of teachers. It is designed to stimulate interest in problems of teacher excellence through presenting new insights developed from recent research. Seven examples are advanced of contemporary thinking on problems faced by teachers in classrooms plus historical and integrative materials necessary to explore the effectiveness problem in broad context.

The introductory chapter presented here has three major emphases. It begins with a conceptual discussion of the basic variables involved in the effectiveness of teachers and some detailed observations on how (and how not) to measure them. It also contains a suggested definition for teacher competence designed to guide further research on this complex topic. Finally, it ends with a summary of the substantive chapters that follow and some recommendations for research to come.

The topic is *research* on teacher effectiveness; not the selecting, training, or evaluating of competency in teachers. Readers must not assume that they will find *The Competent Teacher* neatly identified and packaged within these pages. On the contrary, the problem of teacher effectiveness is so complex that no one today knows what *The Competent Teacher* is. Rather, the concern is with how to study effectiveness as a phenomenon, with concepts appropriate to that task, with methods, and with research strategy. Although the educator will find here a tapestry of new ideas about education (he may even discover significant facts about competence), the basic message is for the educational researcher.

The Problem

In its broadest meaning, competence is an individual's ability to produce agreed-upon results. Factory workers are competent when they perform manipulative tasks rapidly and without errors; scientists demonstrate competency by providing an understanding of the world about us. When is a teacher competent?

This question has brought about probably more research than any other in education. As Gage (1960) has noted, not only is the literature on teacher competence overwhelming, but even bibliographies on the subject are becoming unmanageable. Literally thousands of studies have been reported dealing with characteristics of teachers (rated or measured), effects of teaching, goals of education, and other related issues. Yet, few if any "facts" seem to have been established concerning teacher effectiveness, no approved method of measuring competence has been accepted, and no methods of promoting teacher adequacy have been widely adopted.

The question is not an idle one. There is an unprecedented demand for skilled teachers. More teachers are needed than ever before; these must bear the burden of meeting demands for an ever-increasing level of education in both professional careers and citizenship. Teachers are being called upon for higher levels of performance, administrators are expected both to enlarge their schools and to raise the standards of education, teacher educators are flooded with burgeoning enrollments. An increased understanding of teacher effectiveness will help each of these groups and strengthen greatly the position of American education.

Why, then, is so little known about the effectiveness of teachers?

The author suggests two answers: confusion, and the complexity of the problem.

Confusion

There are several reasons for confusion about effectiveness. First, some educators do not recognize the problem of effectiveness at all. For instance, teachers may see performance in the classroom as a highly personal affair, not a fit subject for the vagaries and impersonality of educational research. Some school administrators, in contrast, convinced of their ability to judge teacher competence, see no reason for research on the subject. School boards often are concerned primarily with finding *enough* teachers—competent or incompetent. Even teacher educators are guilty of ignoring the problem, particularly when they propose new programs whose effects are too "obvious" to require documentation.

To each of these groups must be repeated what the majority of educational researchers have been saying for the last decade: *we do not know how to define, prepare for, or measure teacher competence.* The bulk of studies on teacher effectiveness to date have produced negligible results. Further, until a great deal more is known about classroom interaction, the bulk of educational theories must be judged "untested."

Secondly, there is disagreement over the effects a teacher is called upon to produce. For example, should a teacher's tasks be defined in terms of the ultimate goals of education or in terms of a specific effect upon the pupil? Should a teacher be called upon for equal competence with all types of pupils, or should special competencies be allowed in working with the underprivileged, the handicapped, the exceptional pupil? Are teachers to be judged in terms of the aims of "classical" or "progressive" education? Barr (1950) has suggested that often the teacher is thought of simultaneously,

> (a) as a director of learning, (b) as a friend and counselor of pupils, (c) as a member of a group of professional persons, and (d) as a citizen participating in various community activities.

Are teachers to be judged for competence against each of these simultaneously? *Until effects desired of the teacher are decided upon, no adequate definition of teacher competence is possible.*

Finally, confusion has resulted from using a variety of terms for disparate purposes. Competence, for example, has been used to refer to training processes, properties of teachers, behaviors exhibited by teachers, and effects produced by teachers. The same variables have been termed (by other investigators) "effectiveness," "criteria for competence," "ability to teach," and a host of other terms. *Research on teacher competence becomes effective only with an agreement upon language and the variables for which words stand.* In this chapter, "teacher competence" refers to the ability to produce agreed-upon effects—a teacher property. The term, "teacher effectiveness," points to the effects of a teacher in a given classroom situation.

Complexity

The other difficulty met is the complexity of the problem. Teacher effects, for instance, are often nearly indistinguishable from the effects of other teachers, other agents (such as the parents), or alternate situations. Unlike the factory worker, the teacher does not operate upon the pupil in isolation from other agents. The long-term effects of a teacher are difficult to assess because of the problem of isolating her contributions from those of others, and often the immediate effects of a teacher may be unobservable or have little meaning in relation to educational goals.

Even if the effects of a single teacher were to be isolated, the problem is still complex. Take, for example, a single teacher trait—*warmth*. It may be assumed that *warm* teachers have an effect upon pupils that is different from the effect produced by *indifferent* or *cold* teachers. But is this universally true? It has been suggested that warmth has a positive effect on first graders, but is it equally effective for sixth graders or high school students? Does warmth operate similarly in upper- and lower-class neighborhoods, with members of ethnic minorities, or with emotionally disturbed children? Is warmth equally necessary in the classroom, the study hall, the library, or the gym; should warmth be applied in the control of deviancy or the maintenance of order; and is warmth necessary to the same degree in young and old teachers, in men and women? Such questions have not been answered, nor is it certain that they are the right questions. However, it is certain that competence involves a complex interaction between teacher properties and contextual factors in the community, school, and classroom. There may even be a number of quite independent competencies found in distinct types of teachers. If so, the competence problem becomes one of fitting the right teacher to the right job, and rating teachers on a single scale of competency is highly questionable.

A Seven-Variable Model for Teacher Effectiveness

A primary aim in this initial chapter is the clarification of the effectiveness problem. The problem is approached by offering a model for variables involved in teacher effectiveness research. The model is then explored for its methodological and practical implications and is finally used to summarize the other chapters in the volume.

Chapter 2 is also an introductory contribution and deals with the historical development of the teacher evaluation movement.

Beginning with Chapter 3, substantive papers deal with new and potentially fruitful ways of thinking about teaching. Each of these chapters summarizes a program of research that has been under way for several years; each represents a departure from traditional research on teacher competence; and each suggests new insights on concepts, methods, and (occasionally) facts pertaining to teacher effectiveness. These papers utilize a wide variety of concepts in examining classroom interaction and characteristics of the teacher. Despite this variety, it is possible to collect types of variables together in a simple model.

It has been said that the central problem in understanding teacher effectiveness is establishing relationships between teacher behaviors and teacher effects. The statement suggests that two classes of variables are minimally necessary in the study of effectiveness: *teacher behaviors* (an independent variable) and *teacher effects* (a dependent variable). The problem becomes complex because teacher-pupil interaction is imbedded in historical, social, and physical contexts that constrain and interact with it.

It is additionally compounded when examining both short- and long-range effects and when considering types of teachers or teacher properties rather than the specifics of teacher behavior.

Such considerations suggest a variable system composed of seven classes. To differentiate:

Formative experiences	Main sequence variables
Teacher properties	
Teacher behaviors	
Immediate effects	
Long-term consequences	
Classroom situations	Contextual variables
School and community contexts	

These variables are represented visually in Figure 1 in terms of a system of postulated, cause-and-effect relationships. More explicitly, five of the above variables are postulated to form a cause-and-effect sequence.

1. *Formative experiences, teacher properties, teacher behaviors, immediate effects*, and *long-term consequences* form a sequence such that each variable class in the sequence causes effects in the next variable class listed.

The other two variables are postulated to be contexts for portions of the main sequence.

2. The *classroom situation* imbeds (and interacts) with *teacher properties, teacher behaviors*, and *immediate effects*.
3. *School and community contexts* imbed (and interact) with *formative experiences, teacher properties, teacher behaviors, immediate effects*, and *long-term consequences*.

For the reader such broad propositions may have more sound than substance. It is useful to consider one or two examples in the forms of hypotheses before turning to detailed examinations of each variable class. For ease in comparison, each hypothesis is numbered in terms of the above postulates.

Hypothesis 1 *a.* Teachers receiving "four years of college training" will "know more about the techniques of elementary education" than those receiving less education.

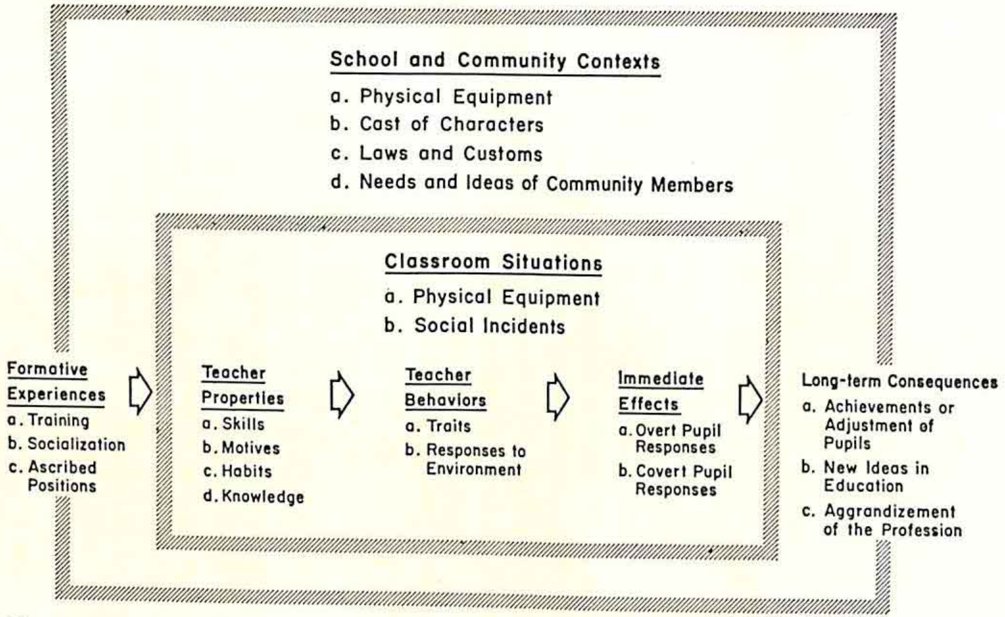


Fig. 1. A seven-variable-class model for teacher effectiveness. Note: *Variables listed in each class are examples.*

b. Teachers "knowing more about the techniques of elementary education" will use more "flexibility in the control of classroom discipline" than those who know less.

c. Teachers who use more "flexibility in the control of classroom discipline" will produce fewer "overt acts of deviancy" by pupils than those who use less.

d. Pupils who exhibit fewer "overt acts of deviancy" will show greater "achievement" at the end of their schooling than those who exhibit more.

Hypothesis 2 "Deviancy control" is more related to "achievement" in classroom situations characterized by "ritual and boredom" than in classroom situations.

Hypothesis 3 The "control of deviancy" is more related to "achievement" in "lower class" schools than in "upper class" schools.

In the above hypotheses "years of college training" belongs to the variable class *formative experiences*, "knowledge about the techniques of elementary education" is a *teacher property*, and so on. Please note that each hypothesis is merely an example of the more general propositions stated above. None of these hypotheses is necessarily confirmable, by the way, but each may be tested experimentally.

Formative Experiences

Formative experiences for the teacher include everything she has encountered prior to her classroom performance and, by extension, the unique ways she has been treated because of her sex, ethnic membership, and so on. Certain variables, however, seem more likely to affect teacher effectiveness than others (and have been more often studied to date).

The most obvious example of these is teacher training. It has been assumed that additional hours of education add up to a more competent teacher. Investigators, of course, also have examined the effects of a major in education, fifth-year plans, in-service training programs, teachers' college versus arts and science background, plus a variety of specific training programs.

Less often examined, but still of potential importance, are prior teacher experiences in terms of residence (size or location of area in which the teacher was reared), family, religion, public or private school experience, prior relationships with teachers, and so on.

Some of the most important formative experiences a teacher has are associated with the ascribed positions appropriate to her. Teachers may be classified in terms of sex, age, physical attributes (such as weight or voice quality), status of infirmity, membership in ethnic groups, social class, nationality of origin, or marital status. Ascribed positions may be based on "unchangeable" physical characteristics (sex, age, or race) or upon one's parents' national origins or social class membership. All ascribed positions subject the person to a unique constellation of experiences that are both historical and contemporaneous.

It may be true, for instance, that the Negro teacher has suffered from a variety of environmental deprivations and may be expected to behave in school situations somewhat differently from her Caucasian counterpart. It is also true that pupils, administrators, and parents will react differently to the Negro (or Mexican, or aged, or married) teacher. Often it is difficult to separate the effects of formative experiences from those of contemporaneous treatment accorded to teachers of certain ascribed positions. However, as Peterson points out (in Chapter 9), all teachers do hold membership in ascribed positions, and the effects of position membership are coercive—whether expected or unexpected.

Each of the above formative experiences "may" affect the competence of the teacher. The exact conditions under which each experience will relate to later effectiveness have not yet been explored in a systematic fashion.

Teacher Properties

A legion of psychological traits, motives, abilities, or attitudes are said to relate to the competence of teachers. Such properties have two features in common: they are hypothetical constructs in psychology, thus they are presumed to characterize the individual teacher in a consistent fashion, over time, and serve to explain her behavior in response to a variety of situations. It is also presumed that such properties are laid "within" the teacher and are not amenable to direct observation in the same way that behavior can be observed. Contemporary American ideology stresses the alterability of teacher properties. We know that warmth, authoritarianism, hostility—and even intelligence and physical aptitudes—are influenced by early learning and can be altered through appropriate educational experiences (including psychotherapy) in later life.

Examples are in order. It has been claimed by various experts that teacher competence requires a certain intelligence level. Other investigators have reported that a warm, firm, democratic, or empathic teacher is the most competent. Each of these qualities is presumed to characterize certain teachers (and not others), to be to some extent central to the personalities of the teachers involved, and to account for patterns of classroom behavior exhibited by them. Probably the most widely studied properties of the teacher are her expertise and knowledge about educational matters.

However, one cannot observe intelligence, warmth, knowledge, or empathy directly. One *can* observe the putative effects of such qualities in teacher behavior; one can ask teachers (or others) to report or rate such qualities, or one can provide objective tests that attempt to measure such properites; one can assume that such properties are produced by certain formative experiences. None of these methods, however, presumes to tap teacher properties directly.

There are several difficulties with the teacher-property concept. The lack of an agreed-upon method for measurement has led to fruitless controversy as to which method is "really" measuring the property in question. Even for such a well-established property as intelligence there remain discouragingly low correlations among the most accepted measurement techniques. Finally, despite years of effort by dozens of investigators, almost every teacher property investigated has shown sharply contradictory relationships with competence.

These problems have led some investigators to advocate doing away with concepts on the level of teacher properties. Thus, Skinner (1950) suggests it is sufficient to establish predictable patterns of response (in this case, teacher behaviors) and that it is only confusing to attempt to look behind those patterns for underlying psychological properties of the person. Others—including the author—would continue to advocate the utility of the teacher-property concept. Phenomenologists argue forcefully that more can be predicted about the behavior of someone by understanding how the world looks to him; and it is difficult to understand archaic, mistaken, or neurotic behavior without concepts of personality that tie together the person's present behavior with his previous experiences.

Pains should be taken, however, to use a vocabulary for teacher properties distinct from that used to describe patterns of teacher behavior. Too often observation of repeated behavior leads to the conclusion that the teacher has a "habit" or is "motivated" to such performance. As Gump points out (in Chapter 6), often the behavior pattern is forced into being by the setting in which the behavior occurs—and the personality of the teacher has little effect. The author also argues (with Turner, Flanders, and Meux and Smith) that far too little effort has been put into observing what teachers actually do, and that relationships between teacher properties and teacher behavior depend upon a classroom, school, and community context.

Teacher Behaviors

From the point of view of the pupil, teacher behaviors form part of the total classroom situation. The teacher, however, views her behavior as a product of interaction between situation demands and personal factors (such as educational philosophy, needs, beliefs, values, and motivations). Regardless of the point of view, teacher behaviors may be observed directly and studied objectively. Unfortunately, few teachers are ever actually observed, either by educational researchers or by principals.

There are endless dimensions of behavior to be observed. Allport and Odbert (1936) have pointed out there are more than 18,000 adjectives available in the English language to describe behavior directly, and a large proportion of these may be applied to the behavior of the teacher. How does one select the behavioral characteristics that are important to observe?

The traditional approach to the study of competence is the selection of general dimensions or "traits" of teacher behavior that are hypothesized (in rare instances, found) to produce given effects in pupils. Perhaps the most sophisticated of all such traditional studies is the one reported by Ryans (Chapter 3).

In three of the chapters in this volume are alternate approaches that seek to classify behaviors in terms of the interactive situation. Turner (Chapter 4) reports a classification based upon consideration of teacher behavior as problem-solving activity. Meux and Smith suggest (Chapter 5) that teacher behavior is episodic in character, and they present a classification of episodes. Flanders (Chapter 7) reports a classification based upon teacher-pupil interaction. In each of these chapters the classification of teacher behaviors is made in terms of situational relevance, thus suggesting that the competent teacher is one who responds appropriately to situational factors.

Once again, however, a *general* classification of teacher behaviors appropriate to the study of effectiveness has not been advanced—nor does it seem likely that a satisfactory system will be produced in the next decade.

Immediate Effects

Unlike long-term consequences, the immediate effects of teacher behavior are played out in the classroom or school. For the most part, investigators of immediate effects have concentrated on pupil response to the teacher, including both overt and covert pupil reactions.

Just as teacher behaviors are a part of the classroom situation for the pupil, pupil behaviors form part of the classroom situation for the teacher. In fact, pupil behaviors must be considered a major component of the classroom situation; and teachers, inadvertently, may cause unwanted situations to arise through mistimed or accidental acts. In this sense, classroom interaction is a total system of interrelated parts; and each act in the system (whether by pupil or teacher) may be seen to have determinants and results in other acts of the system.

This point of view is pessimistic when looking for teacher competence and ignores the greater power and responsibility of the teacher. The fact is that, in most classrooms, the teacher initiates a larger percentage of interactive sequences than do individual pupils. She controls the content of the discussion, operates to keep order and effective response within bounds, and otherwise "takes charge." This suggests that it is possible to examine the immediate effects of teacher activity in the responses of pupils with some assurance that cause-and-effect moves from teacher to pupils.

How, then, are we to decide which qualities of pupil response to measure as indicators of teacher effectiveness? Two viewpoints have prevailed in choosing such pupil indicators. Some investigators have focused upon those qualities of pupil behavior that occur with high frequency and that can be readily observed—such as rates of deviancy, facial expressions, amount of interaction, and the like. The emphasis in such studies is upon immediate pupil response to a single behavioral act performed by the teacher, and often many hundreds of ratings of pupil response may be obtained within the span of an hour's classroom interaction.

Other investigators have chosen pupil qualities that reflect educational goals measuring pupil properties rather than "superficial" behaviors. Studies have been made, for example, of pupil learning, attitudes, sociometric preference, clinical disturbance, and so on. Unfortunately, such measurements are much more difficult than observing pupil behavior and generally are made after the pupil has been exposed to a selected teacher for several hours, days, or years.

From the viewpoint of the behavioral investigator, measurements of pupil properties represent a complex summation of countless contexts, situations, and interactions between teacher and pupils, with the basic rules of summation as yet unknown. From the viewpoint of the pupil-property investigator, measurements of pupil behaviors bear little relationship to the purposes of education. Actually, the two types of studies are complementary, and relationships between pupil behaviors and properties will presumably be nailed down in the not-too-distant future.

However, if teacher effectiveness is assumed to be bounded by context and situation, it is also reasonable to assume that pupil response is unique to these conditions. This suggests that, until such time as there is better information about the relationships between pupil behaviors and properties, focusing upon the former is a better bet. To date, little attention has been paid to the many varieties of pupil behavior observable in response to teacher behavior.

Long-term Consequences

Eventually, of course, it will be desirable to know the long-term consequences of teacher activity on pupil, school, and community.

Those who study teacher effectiveness have the broad goal of discovering teacher qualities that lead to differences in the lives of pupils, whether these differences are reflected in professional achievement, adjustment to the vicissitudes of life, attitudes toward others, or out-and-out financial success. It is intriguing when a noted person attributes his achievements, in part, to a teacher. Far too often (as Peterson points out in Chapter 9), the teacher has but scant feedback of any kind to reward her for years of effort devoted to other people's children.

Of less central importance, perhaps, are the long-term contributions a teacher makes to her school and community. At a minimal level, she must function in a manner that is not disruptive or troublesome. The good teacher can contribute to new ideas in education, to the aggrandizement of the profession, to an understanding of education in the community, and to new roles and functions for education—all within the normally assumed compass of her role. For some teachers, particularly those who move on to administrative positions in the school, such contributions become primary.

Unfortunately, established relationships between teacher activity and such laudable long-term consequences are far off. Once again, to take simply the problem of the teacher's effect upon a pupil, the teacher's impact is inevitably mitigated by the efforts of others; and an understanding of the effects of even one such agent is lacking.

It is wise to keep aware of, indeed to pursue with vigor, the basic goals of education. However, the ability to relate teacher behavior to such goals is probably years in the future.

School and Community Contexts

The societal arena in which classroom education is played out includes physical settings, a cast of characters, established patterns of behavior, and agreed-upon definitions of social events. In general, each of these elements constrains classroom interaction and is not under the teacher's control.

The school usually occupies a building with fixed walls, halls, doors, windows, lighting, heating, and the like. Desks and chairs are provided by the community along with other paraphernalia such as blackboards and chalk, textbooks, audiovisual aids, and—more recently—television and teaching machines. Whether she likes it or not, a teacher must "make do" with these physical arrangements.

Equally constraining is the cast of characters associated with American schools. Teachers exist at the bottom of a status hierarchy that also includes the principal (and his assistants), the superintendent (and his staff), and the school board. The classroom ordinarily contains only the teacher and her pupils, although occasionally there may be a practice teacher, a subject-matter specialist, a visitor. Also to be found in the larger school are librarians, medical and food personnel, janitors, and a variety of additional service employees. In larger schools (particularly at the secondary level) teachers may also find themselves tagged as content specialists; for example, music teacher, coach. In primary schools teachers may be identified with a grade specialty. In addition, the community provides a number of ascribed positions (previously discussed) into which teachers are placed. Standing always in the background are the ubiquitous parents whose opinions (often uninformed) must be considered if the school is to function effectively.

School operation is also largely constrained by law and custom.

The day is organized into classes of a given length, with given titles and assigned curricula. Pupils pass from one subject to the next in a predictable fashion. Teachers are responsible not only for imparting subject matter but also for the maintenance of order and discipline, and a sharp power differential exists even between the weakest teacher and her most obstreperous student. There are major differences among the patterns of behavior allowable in such settings as classrooms, the library, the gym, playgrounds, lavatories, and the teachers' lounge. Additionally, both formal and informal groups may appear among administrators, teachers, and students. Through these groups norms are created and enforced.

The school also bears functional relationships with the larger community. School systems are expected to serve citizen needs ranging from entertainment, to the desire to learn, to the need for status advancement, to baby-sitting. Community expectations may or may not be formalized in terms of explicit goals set before the school. Further, the school interacts with other institutions in the community. It provides employment for a number of people, consumes certain kinds of goods, keeps "wild youth" off the streets, provides an avenue for social service, and so on. The school building is sometimes stockpiled against the threat of tornadoes or nuclear fallout. Often, particularly in small communities, the school may be the only institution attracting the attention of a majority of the citizenry; and observers have noted the demise of some small communities when their schools were consolidated with others. In low population areas a citizen may carry his community status into school settings, such as the PTA. Similarly, school personnel often may find their occupational positions carrying into the small community, such as when teachers are expected to lead in Sunday school or to be paragons of public virtue. Urban schools appear to be more isolated, and professional affiliation there may be easier to shed at the end of the day.

Much of the community-school interplay has yet to be examined for impact on teacher effectiveness. As Peterson points out (in Chapter 9), by a tendency to ignore the effects of age and generational groupings among teachers, competence has been misunderstood. Rosencranz and Biddle suggest (in Chapter 8) that community-held norms have a similarly important effect on effectiveness. There is no investigator known, however, who has suggested a method for differentiating the contextual variables that should, or should not, be investigated for effects on effectiveness.

Classroom Situations

The classroom situation is similar to the school and community context in that it contains both physical and social elements. However, this situation is generally under at least some control by the teacher. Thus, at least a portion of any teacher effectiveness determination must relate to her appropriate manipulation of classroom situations.

Among features of a situation easiest to manipulate is the small-object content of an environment. Teachers everywhere are familiar with the seductive effect of pets, tricky machines, things to build, motion pictures. The effects of teacher control of colors, lighting, distribution of seats, and automated learning are also being explored gradually.

However, situations can also vary in terms of social dimensions. Excluding subject-matter designations (generally not under teacher control), teachers may set up the class as a lecture, a discussion, a seminar, or a free-for-all. There is a world of difference between the participative lesson, the playground romp, the birthday party, the storytelling session, and the study hall.

Situations also can be classified in terms of their microelements. "Pacing" is highly important in the early elementary grades. Gump discusses the "deviancy control incident" (in Chapter 6) and focuses upon teacher response to such a situation and the resulting "ripple effect" among pupils.

It should be possible to provide one or more classification systems for dealing with classroom situations. Barker and Wright (1954) have suggested that both school and community settings "coerce" the behavior of participants. If this observation can be supported in further research, teacher effectiveness must also involve the ability of the teacher to manipulate classroom situations. Thus, teachers might prove competent either because they could set up appropriate situations or because they responded appropriately to naturally occurring situations (or both). As yet, however, no method for classifying classroom situations has proved widely effective.

Recapitulation

A model has been proposed for seven classes of variables involved in teacher effectiveness, together with three postulates stating cause and effect relationships among them. Five main sequence variables were proposed including *formative experiences, teacher properties, teacher behaviors, immediate effects*, and *long-term consequences.* Each of these variables is postulated to affect the one immediately behind it in the sequence. In addition, *classroom situations* (which are somewhat under teacher control) and *school and community contexts* (which are not) jointly constrain and interact with the linear effects of the five main sequence variables listed above.

Numerous variables have been suggested for each variable class, although for no class today is there an effective catalogue of variables likely to be useful in effectiveness research. No attempt has been made to state explicit hypotheses for the model, although a number of examples have been suggested.

The Search for Teacher Competence

Nowhere has the above discussion dealt with teacher effectiveness per se, or its sister concept, teacher competence. Such definitions now may be usefully given.

Let us assume that teacher competence is *one or more abilities of a teacher to produce agreed-upon educational effects.* A number of observations may be made in terms of the variable model:

1. It is necessary to agree upon which educational effects the teacher is to produce; without such agreement no definition of competence is possible. Agreements may be worked out in community discussions or may be imposed (with varying degrees of success) by school boards, administrators, or pressure groups. In the long run, "local" definitions of desired effects must disappear—otherwise pupils can find themselves undereducated, and their teachers incompetent, when they move to a new com-

munity. Many observers have suggested that the responsibility for outlining and encouraging national standards for education lies with educators. The setting of such standards would certainly tend to raise the status of the profession.

2. Teacher competence as an ability belongs within the second variable class, teacher properties. There has not, however, been given a denotative definition of particular abilities involved in competence. Instead, it has been suggested that whichever properties of the teacher produce the agreed-upon effect are the properties constituting competence.
3. Such a definition suggests two empirical research stages bearing directly on competence. In the first stage—when there is little or no idea as to what abilities are involved—competence is undefined unless empirical relationships can be found between teacher abilities (the independent variable) and an agreed-upon educational task (the dependent variable). The latter variable is usually termed the *criterion* for measuring teacher competence. Research during this first stage focuses on those properties of teachers found to affect a given criterion variable.
4. Since competence is an ability, however, it may be possible to develop instruments (similar to IQ or musical ability tests) for objective assessment. Should such an instrument be developed, and its validity demonstrated with acceptable criterion variables in a variety of educational contexts, the strategy for competence research would change. In such a later stage, interest would lie in exploring the interaction between competence (as measured by the instrument) and other variables.
5. Needless to say, there has been no penetration beyond the first stage of research to date. In fact, it has barely been entered.
6. Formative experiences are not directly involved in the definition of teacher competence. On the other hand, those conditions producing competence may be investigated, with competence measured either by a criterion measure or by an instrument designed for the objective assessment of competence abilities. In the former case, however, competence itself is not actually studied; it is, rather, assumed.
7. Considering the probable complex relationships among teacher behaviors, criterion tasks, and contextual variables, it is likely that competence is a cluster of relatively unrelated abilities. Any one of these may be inappropriate to some contexts. For example, some teachers may be inspirational leaders, others warm counselors, and still others walking encyclopedias. In certain contexts, each of these competencies may be highly effective; in others each might have little or negative effect.

The probable existence of two or more teacher competencies useful only in limited contexts have led a number of observers to question the utility of teacher competence as a concept. What should be studied, they argue, is *teacher effectiveness*, or *the ability of a teacher to produce agreed-upon educational effects in a given situation or context.* Like competence, teacher effectiveness is found in the variable class *teacher properties* and each of the first six points above applies equally to the effectiveness concept.

Reasonable as this argument may be, it seems to the author that empirical grounds are not yet established for determining whether there are one or more general abilities adding up to effectiveness in most situations. The problem will be resolved only when an adequate catalogue of educational contexts is provided so as to allow assessment of various teacher abilities for their effects in each.

A second group of observers has recommended abandoning the search for competence on different grounds. Noting the previous unwillingness to study classroom interaction and the poor correlations so far obtained between "direct" measure of competence and criterion measures of pupil growth, they advocate a study of teacher *behaviors* rather than *properties.* The question of competence should arise only after clear evidence of behavioral success.

The author has considerable sympathy with this point of view. However, such advice is of little use to the principal or superintendent under fire to rate teachers for competence. Most administrators are assured that the "best" and "worst" teachers are easy to identify (once more suggesting that some sort of unitary ranking of competence is possible). Thus, despite objections from researchers, competence remains a limited but useful tool for practitioners even though one is not quite sure of its meaning. The researcher ignores the practical problems of competence creation and evaluation at his peril. It is unlikely, however, that significant advances will be made in understanding teacher competence without a clearer picture of teacher behavior and its effects.

The Measurement of Effectiveness Variables

Many different types of measurements have been used for the assessment of effectiveness variables. Techniques have included ratings (by various persons), direct observation of behavior, objective instruments, and other devices. Each of these techniques has been used, moreover, for the measurement of a variety of variables.

A sizable literature also has appeared to deal with the adequacy of various techniques for effectiveness measurement. Unfortunately, much of this literature has confused the measuring technique with the variable being measured. As a result, a measurement may appear in one author's treatment as a "cause" of effectiveness, while another may treat the same measurement as a "direct indication" of effectiveness—or even criterion.

A review of the forms of measurement in use and a discussion of their application to effectiveness variables is pertinent.

Observation Techniques

Ideally, the investigator should gather information about overt variables through the direct observation of behavior. Thus, formative experiences of the teacher should be witnessed, teacher behaviors in the classroom should be recorded mechanically, and all outcomes of education should be traced in terms of the overt responses of pupils and others. In actual practice, of course, the observation of behavior is costly and tedious, and many variables of interest operate internally in the teacher or pupil, or are separated from the classroom scene by the gulf of time. For these reasons and others, most persons doing research on teacher effectiveness have avoided behavioral observation.

Behavioral observation techniques may be classified into four categories, depending upon the method used to record results. *Participant observation* is a technique originally developed by anthropologists and depends upon extensive participation by the observer in typical interactive situations. Thus, the participant observer of classroom interaction would sit in a variety of classrooms "as a pupil" or "as a teacher" in order to develop a feel for what

Acc. No. 3149
Date 12-3-85
370'78
BID
0039
S.C.E.R.T., West Bengal Library

was taking place. Observations are usually written up by the participant observer at the end of the day and are summarized in a journalistic style. In actual practice, most participant observers supplement the direct observation of behavior with intensive depth interviews taken from selected respondents who participate in observed situations. Whenever used, however, participant observation has emphasized the observer's "understanding" of a complex situation rather than the documentation of those "understandings" with firm evidence.

A variety of *categorical check lists* have been suggested for observing classrooms. Such forms emphasize a small number of categories into which the behaviors of teacher or pupils may be organized. Ratings are given once for each interval of time or at the beginning of each type of behavioral emission. Since a limited number of categories for rating behavior must necessarily ignore much of the richness of actual interaction, those listed in a rating form have generally been chosen for a specific purpose. Bales's (1958) categories were chosen to represent a theory of behavioral interaction. Flanders (Chapter 7) chose categories he felt would give the best prediction of teacher effects.

Limitations of the check list have led Barker and Wright (1954) to develop a *specimen record* technique that attempts to get at behavioral richness through the recording of innumerable details of interaction. In using this technique, the observer focuses upon a specific participant (such as a pupil) and records all things that the pupil does or says together with such information about the environment as may be necessary to explain the participant's actions. Records are kept in the form of handwritten notes, and each observer is allowed to retire from the scene after a suitable interval (not more than half an hour) to dictate an extended account of all that happened. The observer is instructed to use low-level abstractions in describing behavior and to avoid interpretive remarks except for short explanations to clear up possibly confusing sequences. Specimen records thus obtained are later coded for events and their dimensions so as to discover the outstanding features of those events. Gump (in Chapter 6) reports a number of studies using the specimen record.

Finally, the development of new devices has allowed for the *electronic recording* of behavior. Thus, sound-tape recordings of classroom interaction have been made by a number of investigators. Others are exploring the use of kinescopic recordings, motion pictures, and TV tape recordings for a visual record. The advantage of such recordings is that the behavioral sequence can be played back over and over in order to pick up subtleties that would have been missed by using cruder techniques. Of course, the actual coding of interaction necessitates use of either a check list or the assembly of a *specimen record.* Meux and Smith report (in Chapter 5) the use of electronic records combined with the specimen-record technique.

To summarize, participant observation is best used when one is totally unfamiliar with the situation and when one wants an overview in order to develop hypotheses. It suffers from a lack of "hard" evidence. Categorical check lists allow the collection of considerable data with high reliability but tend to miss subtle details of interaction because of the limited categories used. Specimen records are more expensive than rating forms but allow a more intensive study of the events surrounding and emitted by a subject. Their reliability may be questioned unless records are taken from electronic recordings. This last technique is by all means the most expensive, but it is highly reliable and allows for the greatest flexibility in converting to categorical records.

It is likely that greater use will be made of behavioral observation during the next decade. Indeed, it may be said that progress in understanding teacher competence depends upon such methods. Behavioral observation is certainly the outstanding technique for studying teacher behavior, immediate pupil response, and classroom situation. For these classes of variables it seems to be irreplaceable.

Objective Instruments

Objective instruments are similarly necessary for the assessment of covert variables in teacher, pupil, or others. Provided that one can assume an honest, motivated, self-conscious respondent, any conceivable psychological property can be assessed by designing the proper instrument.

These assumptions may often be suspect. Teachers may not be aware of their motivations or may hide them with cloaks of dishonesty or rationalization. Accurate assessment of pupil learning may be hindered by the pupil's misunderstanding of the task, by fatigue, by lack of motivation. It may not always be possible to present the respondent with contextual or "real" stimuli to bring forth the constellation of properties to be measured.

Despite these and other shortcomings, the ability of human respondents to describe their phenomenal experiences provides a unique "window"—opened best with objective instruments—into covert processes. There are many types of objective instruments, and all of the following have been used in competence research.

Achievement tests are familiar to the American school child, as they are the universal and most used method of measuring classroom learning. Such tests are usually composed of a series of graded tasks with one or more scores assembled to provide a profile of achievement. Usually achievement tests are developed with great care and, as has been previously discussed, must be validated against a predicted criterion of pupil behavior. They may, obviously, be used also for the assessment of teacher achievement (following, for instance, a refresher course). This is less often done.

Ability inventories are similar in form to achievement tests and are similarly designed for self-administration by subjects. They differ from achievement tests in that abilities are presumed to be unchanging properties of the person or to change as some predictable feature of growth. An IQ test, for example, is basically an achievement test. It becomes an ability inventory only when one assumes a constant ratio between achievement age and chronological age in computing the IQ itself. Most tests of musical ability assume "inherited" or unchanging traits. Ability inventories are often used today to assess neurotic tendencies or other properties that would influence job success in business firms. As yet, they are not often found in school systems.

Questionnaires and *interview schedules* differ from tests and inventories in that they are designed primarily for research purposes. As a result, they ofen produce as many different scores as there are questions. Questionnaires are designed for self-administration by respondents who are often at a distance from the investigator. Interview schedules are designed for administration by an interviewer who asks questions of the respondent and records answers for him. In general, questionnaires are cheaper and are recommended when answers may be precoded into easily understood categories. Interviewing is a better technique when it is desired that a topic be pursued in depth, when precoding of answers is difficult or impossible, or (obviously) when interviewing illiterate respondents such as preschool children. It is also necessary to interview certain groups who respond poorly to questionnaires such as very low income groups, physicians, or businessmen. Literally thousands of questionnaires and interview schedules have been developed for research on teacher competence using varied respondents including teachers, pupils, parents, administrators, school board members, and school janitors.

Projective tests are a form of interview schedule designed to tap "unconscious," or hidden, properties of the respondent. In the normal interview schedule the respondent is made aware of the purpose of the question so that he can reply as honestly and fully as possible. In the projective test the purpose is hidden from him and he "projects" motives, beliefs, values, and other properties into his answer. Because of administrative difficulties and problems of moral kickback, projective tests are almost always administered. Common projective tests (such as the Rorschach and TAT) have been used for research on teacher competence, while recently several projective tests have been developed specifically for investigation of teacher motivations and the like.

Regardless of the objective instrument chosen, it is designed for the tapping of covert qualities. Consequently, these must be designated the primary method for measuring teacher properties and those covert responses of pupils (and others) considered to be outcomes of teacher activity. Once again, however, instruments designed for measuring covert qualities must be validated against criteria of acceptable behaviors *in situ.*

Rating Forms

Probably the most common technique used in research on teacher effectiveness is the rating form. In a typical application, one is asked to rate the teacher's ability, the teacher's performance, or pupil response to the teacher. Such ratings have been used as a criterion for evaluating teacher competence or predicting "to the outcome" variables in the classroom—which were then measured objectively. A fantastic cast of persons has been used as raters in competence research. The group includes the classroom teacher herself, other teachers, teachers-in-training, critic teachers, principals, supervisors, superintendents, school board members, pupils, parents or lay persons, college instructors, and others. The vast majority of studies reported on teacher competence have made use of rating forms for one purpose or another.

Generally, the results of research using rating forms have been poor and contradictory. This is not surprising in view of the fact that such judges as listed above are handicapped by personal bias, a lack of training for observation, and a lack of firsthand information concerning the teacher or classroom interaction. Yet each year brings a new crop of studies using rating forms. Why this perseverence? One obvious answer is the prevalence of rating forms in school programs of assessment, merit pay, and promotion. Davis points out (in Chapter 2) that the practice of evaluating teacher competence through the use of rating forms goes back many years and is related to ideology and to similar movements in government and industry. Pressure is felt in many American communities to promote and raise the pay of only the best teachers. To those who are strongly committed to "quality" education, even a poor program of teacher evaluation seems better than no program at all. Probably the majority of educational administrators feel (with some justification) that they can identify the "best" and "worst" teachers. Many teachers themselves enjoy working in schools that provide "feedback" through the regular use of rating forms. Thus, despite doubt-casting evidence, there are strong pressures to keep rating instruments in the schools.

As long as rating forms are being used, there will be a temptation to utilize the data they provide for research purposes and to hope that significant knowledge about the competence of teachers will be forthcoming. This temptation appears strongest for the doctoral candidate who attempts to complete his dissertation at a distance from the campus and who is familiar with the problems and challenges of the rating system used in his own school. Ratings, unfortunately, offer the investigator an unexamined hodgepodge of classroom relationships and confusions stemming from the rater's biases and lack of information. It is quite possible, of course, that ratings by a trained and unbiased person sometimes may be valid. However, until careful studies are made of the facts and artifacts involved in the rating process, ratings seem less than useful for research on teacher effectiveness.

Self-Reports

Self-reports relate both to objective instruments and rating forms. In self-reports, however, the respondent is asked to provide information from his own experiences—to replace the direct observation of past events with his own recollections. Self-reports differ from the usual objective instrument in that psychological properties of the subject are not being studied. They are, instead, relied upon to replace objective evidence not available. Self-reports differ from rating forms in that categorical judgments are not requested. Rather, the respondent is asked to describe what actually happened. As with questionnaires and interview schedules, the self-report may be given in open-end or "depth" statements, or it may be answered as a series of multiple-choice alternatives to predesigned questions.

The use of self-reports has a long history in sociology; indeed, it constitutes (along with use of extant records) a major source of data for the empirical sociologist. Self-records are likely to be used for information relating to formative experiences, classroom situations, and school and community contexts. Self-reports are vitally important in the assessment of formative experiences, and they are the best source of evidence for broad contextual variables. The use of self-reports for study of classroom situations should be questioned, however. Such reports inevitably reflect the cognitive categories used by the respondent for thinking about the subject matter he is asked to describe. It seems likely that adequate categories for describing classroom situations are not prominent in the cognitive fields of such respondents as teachers or pupils. (Indeed, if an adequate categorical system is known to an educational researcher, he has not yet made his insight available to others.) Asking a teacher to describe classroom situations is likely to cause her to repeat educational ideology, while parents and pupils will report recent pleasant or unpleasant classroom experiences.

There have been few explorations of respondent biases in the use of self-reports. This technique is, therefore, useful only when objective data are impossible or expensive to get.

Existing Records

In the general archives is a variety of records useful in research on competence. These range, in generality, from the U.S. census, to state records of school expenditures and enrollment, to local records maintained by the school board, system, or school. They are particularly useful in documenting changes that have taken place in the school system as a result of historical pressures—thus helping to establish the school and community context. They may also provide direct evidence of formative experiences, as when one may trace, through records, the effects of a given teacher, school, or community in the later performance of teachers or citizens.

Unfortunately, little use has been made of the majority of existing records in teacher effectiveness research. The major reason for this is that school records have been thought by most researchers to be of little value in the effectiveness problem. Thus, there are voluminous records of certain variables, on the one hand, and voluminous research projects on classroom interaction, on the other, with little intercourse between them. A number of variables should be examined in existing records for their impact on teacher role and effectiveness. For instance, Peterson (Chapter 9) suggests the importance of age, sex, and generational differences on the satisfactions and effectiveness of teachers. Peterson's observations might be tested through the use of school records and by comparing several types of communities.

It should be emphasized, however, that most existing records are inadequate for research on effectiveness variables. Not only are they often kept on unimportant subjects (such as number of pieces of chalk used), but they also may be seriously in error due to administrative indolence or falsification. What school, for instance, has a careful, ten-year file of interaction records taken from the classroom or a similar series of objective test results administered to all teachers? A simple, inexpensive, but enormously significant contribution to research on teacher effects could be made by the school administrator who is cooperating with a team of educational researchers to collect a variety of agreed-upon observation, rating, report, and test results for one or two decades. Such a project is providing vital results in the study of personality (Kagan and Moss, 1962) and genius (Terman and Oden, 1959). It should have an equally powerful effect on understanding the effects of teaching.

A Priori Classification

Finally, some variables in effectiveness research are amenable to a priori classification by the investigator. In the usual field study, such variables are likely to be ascribed positions to which the teacher belongs (such as sex or racial classification) or contextual variables such as school size, characteristics of the community, and the like. In the rare experimental studies of teacher effectiveness, such elements as manipulations of classroom situation, pupil response, or teacher technique manipulated by the experimenter tend to fall into a priori classifications provided by him. In all cases, measurement by classification is sufficiently obvious as to render useless any more complex method of measurement, although the utility of any a priori classification must be justified in terms of results on the effectiveness problem it produces.

Summarizing Methods

Salient features of methodology in effectiveness research may be seen in Figure 2, which is organized around a recommended set of methods for each variable class. As a rule, measurements by a priori classification, behavioral observation, and objective instruments are to be advocated over measurements made by existing records, self-reports, and (above all) rating. These also are advocated: (1) the use of electronic recordings of classroom inter-

	Competence Variables						
	Formative Experiences	Teacher Properties	Teacher Behaviors	Immediate Effects	Long-term Consequences	School and Community Contexts	Classroom Situations
Recommended Measuring Techniques	Existing Records A Priori Classification	Objective Instruments	Behavioral Observation	Behavioral Observation Objective Instruments	Behavioral Observation Objective Instruments	Existing Records Objective Instruments	Behavioral Observation
Alternate Measuring Techniques	Self Reports	Self Reports Behavioral Observation	Self Reports	Self Reports	Self Reports	Self Reports A Priori Classification	Self Reports A Priori Classification
Poor Measuring Techniques	Rating Forms						

Fig. 2. Techniques recommended for measuring teacher effectiveness variables.

action; (2) the elimination of rating forms in serious research on teacher effectiveness until an understanding of their biases is available; (3) the greater use of existing records; (4) the extending of school record keeping to include behavioral observation and objective instrument records, in depth and over a period of decades.

Integrating the Contributions

A short description of each substantive contribution to the volume follows. (Chapter 2 by Davis is a history of the competence-evaluation movement and, being vitally significant, is reviewed here.)

The Teacher Characteristics Study (Chapter 3)

The study described by Ryans in Chapter 3 is both the most classical and the broadest presented in the volume. Ryans' major focus is on "teacher characteristics" by which he means both teacher properties and teacher behavior. He presents three dimensions of teacher behavior (measured by rating forms used in direct behavioral observation): *warmth, organization,* and *stimulation.* He also details seven teacher characteristics (measured by objective instruments) including: *favorable opinion of pupils, favorable opinion of classroom procedures, favorable opinion of personnel, traditional versus child-centered approach, verbal understanding, emotional stability,* and *validity of response.* A portion of Ryans' paper is devoted to explaining these ten measures, their development and measurement. He then presents a variety of results showing relationships among the measures as well as their relationship to other variables. The latter include formative experiences (sex, age, and early activities), the context of school size, and pupil behavior in the classroom. Each of these variables except the last was measured by self-reports, while pupil behavior was rated from direct behavioral observation. Ryans reports additional relationships concerning his central ten variables in other published reports.

Ryans' design is classical in the sense that teacher "characteristics" are abstracted from the classroom context. Thus, classroom situation and teacher-pupil interaction are ignored for the most part. At the same time, Ryans' efforts have been characterized by methodological sophistication and breadth of variables considered. His work is unique in that he has established relationships between the ten "characteristics" described and both formative and outcome variables. Other researchers would do well to begin their more complexly designed programs of research with the assumption that Ryans' general variables are of significance.

Teaching as Problem-solving Behavior (Chapter 4)

In Chapter 4, Turner details an explicit strategy for effectiveness research focused upon the assumption that teaching may be viewed as a series of problem-solving or coping behaviors. Utilizing this strategy, Turner and Fattu have developed objective instruments to measure potentialities in teachers for performing coping tasks. Turner then shows that teacher scores on these instruments are related to formative experiences, other teacher properties (such as intelligence and attitudes or values), and to such contextual variables as subject matter and age of pupils. Turner also reports additional relationships between his task variables and others in additional published reports.

The design followed by Turner is intermediate between the classical approach of Ryans and the systematic interaction studies of Meux and Smith, or Flanders. Turner assumes that the teacher is reacting to problems posed by the classroom situation; thus he builds classroom contexts into the definition of teacher properties to be measured. By way of contrast, Ryans' ten traits are abstracted above the level of the classroom. It is not clear at this stage of the research program whether or not Turner is correct in his assumption about the coping behavior of teachers—nor does he attempt to document this assumption directly with observation of teacher behaviors. If Turner is correct, in the next five years there will be a variety of results relating teacher abilities (as measured by Turner's instruments) and criterion variables, such as pupil learning.

The Logic of Teaching (Chapter 5)

Meux and Smith make, in Chapter 5, an assumption similar to Turner's; that is, teaching behavior may be classified in terms of its logical qualities. However, instead of assuming that teachers have abilities measurable for logical response, Meux and Smith report a series of studies based on the direct observation of teachers in classrooms. They prepare sound-tape recordings of cases of classroom interaction, develop specimen records from these original recordings, and then classify classroom interaction into episodes (which may be either *monologues* or *dialogues* of teacher-pupil interaction). Much of their paper is taken up with a classification scheme for dealing with such episodes while, because the project is still in its early stages, few data are reported relating these teacher behavior variables to others about which the authors speculate.

Meux and Smith are unique in this volume because of their detailed treatment of teacher behavior as a variable. Flanders focuses upon the same variable but with a categorical check list. Other categories of teacher behavior can of course be coded (see, for example, Kounin, Gump, and Ryans, 1961), but the work of Meux and Smith focuses singularly on the logic of teacher-pupil interaction. The project appears to have exciting potentialities.

The Manipulation of Classroom Situations (Chapter 6)

In Chapter 6, Gump reports on three programs of research that have in common a concern for environmental variables such as the classroom situation. Reviewed are the work of Kounin and Gump on the "ripple effect" of deviancy control, research by Redl and Gump on children's activities, and the series of studies by Barker and others on the psychological ecology of a small, midwestern community (including its school). With a few exceptions, such studies have made use of direct behavioral observation and the specimen-record technique. However, as opposed to the work of Meux and Smith, emphasis has been upon the objective description of situations and pupil response rather than on teacher behavior. Gump reviews evidence for the proposition that classroom situations coerce the behaviors of both teacher and pupil, and that management of the situation is a major ingredient of teacher competence. He also discusses relationships between classroom situations, such molar contextual variables as school size, and outcomes for both pupils and other school participants.

Although similar to Meux and Smith in the use of specimen-record techniques, the studies reported by Gump are unique in this volume in their concern with the immediate environment of learning. Such concepts as Barker's *setting* and Redl's *seductive prop* have strong heuristic value leading toward the as-yet-missing general theory of classroom situations. Again, the studies reported in this chapter are young and may be expected to be of much value in understanding teacher effectiveness in the next decade.

Interaction Analysis in the Classroom (Chapter 7)

Chapter 7 by Flanders presents an outspoken defense of behavioral observation combined with the use of a categorical check list. Flanders has developed a form combining seven categories of teacher behavior with three of pupil response. He presents evidence to show that different patterns of interaction obtain for teachers of varying abilities, subject matters, and national contexts. Utilizing an interesting method of analysis involving pairs of interaction incidents, Flanders is able to show that the immediately preceding pupil behavior is a situation to which the teacher is responding, while pupil behaviors serve as the major criterion variable for examining the effects of teacher behavior. Flanders presents numerous findings in table form and refers the reader to other published studies in which additional evidence may be reviewed. These include findings on the retraining of classroom teachers for sophistication in their classroom techniques.

Flanders' work represents a simple yet powerful method for studying teacher behavior and should provide a wealth of evidence relating to matters of competence. If Flanders is correct in his observation, the usual traits studied in classical research on teacher competence are of little value in predicting classroom behavior—which must stand midway between teacher properties and pupil responses. This suggests that even such sophisticated studies as Ryans' and Turner's will ultimately produce little because of their inability to handle the obviously important variables of teacher behavior. Perhaps the usual ideology taught in schools of education bears little resemblance to the actual situations teachers face and their behavioral responses. One may question Flanders' restriction of coverage in the seven categories of teacher behavior chosen; one cannot but be impressed with the strength of his findings.

The Study of Teacher Role (Chapter 8)

Chapter 8 by Rosencranz and Biddle is a maverick in several senses. The authors are less concerned about classroom interaction than any others in the volume. Emphasis is on a complex model of cognitions held about the role of the teacher by teachers, administrators, parents, pupils, and others in the community. In fact, the model used is more complex than the seven-variable class model suggested for *this* chapter. The authors are concerned with *expectations* (or anticipations) and *norms* (or values) involving the teacher, and make the point that *role conflict, pluralistic ignorance,* and other role phenomena form part of the context with which all teachers must deal. These determine the teacher's ability to survive and produce in school and nonschool contexts. Findings reported by Rosencranz and Biddle are based upon extensive field studies using group-administered questionnaires given to respondents in a variety of social positions. The authors also make a plea for acceptance of role theory as an orientation best expressing problems that center around the effectiveness question.

The chapters by Rosencranz and Biddle as well as Peterson (Chapter 9) are alone, in this volume, in their attempt to handle in depth the more molar contexts of school and community. In reference to the school, contexts are to be found in shared cognitions existing (or stimulating controversy) among various members of the educational community. Thus, they see the teacher's task in the classroom as one of fulfilling a multiplicity of demands placed upon her by various participants in the system. If Rosencranz and Biddle are correct, one can predict which classroom situations will produce learning, and which teacher behaviors will be effective, through a knowledge of the cognitive fields of teachers, pupils, and others. Thus, the authors' theoretical position also will underlie such interactive sequences as can be observed in Flanders' data and are suggested by Meux and Smith, and others.

Age, Sex, and Generation among Teachers (Chapter 9)

In the last chapter, Peterson turns to the ascriptive categories characterizing teachers and the problems and effects on competence these engender. By means of intensive interviews, Peterson has assembled insights and propositions relating to age and age roles, generational cliques, historical changes, and problems of recognition, involvement, and identity for teachers in the public school.

Peterson, to a greater degree than the other authors, deals with the large contexts in which classroom education takes place in America. Although he works in depth only with ascriptive categories, he succeeds in focusing attention on the plight of the schoolteacher—as opposed to others who tend to divorce behaviors, cognitions, or abilities from the teachers who exhibit them. Once again, the problem is not one of effectiveness directly, but rather of understanding the basic conditions under which effectiveness is played out, the basic rules stemming from our society and educational institutions constraining activity choices for teacher, administrator, and pupil. If Peterson is right, quite different patterns of effectiveness must characterize the young teacher, the old, the man, the woman, the white, the Negro. Perhaps such variables are ignored in other chapters (except Chapter 8) only because of the dedication of other authors to their own research models. It must be noted that, despite the obvious importance of such contextual variables as are discussed by Peterson, no one yet knows for certain how they interact with the more focused variables of classroom interaction.

Summary and Recommendations

It has become evident by now that several common threads run through the majority of contributions presented. These threads should be examined in terms of future research on the competence problem.

The authors of the substantive chapters without exception are engaged in long-term research projects in which activities are planned over a period of years rather than months. These projects are characterized by unique concepts developed for purposes not necessarily focused upon the competence problem itself. Thus, each project is likely to "pay off" in a variety of practical ways of which the effectiveness problem is but one.

Such projects are unlike the types of research commonly engaged in by doctoral candidates and supported by such federal agencies as the Cooperative Research Program of the Office of Education or the research grants of the National Institute of Mental Health. It seems likely that little of significance will come from dissertation research unless the dissertations are imbedded in broad programs as described here. Strongly recommended with regard to federal support of research are long-term commitments for basic research on the development of concepts, methods, and propositional structures. It is possible that national or regional education associations might also take the lead in providing commitments for long-term support of basic research.

A second observation deals with the necessity of developing concepts and data concerning teacher-pupil interaction. Four of the substantive papers focus directly on the problems of behavioral interaction in the classroom, and there is general agreement that adequate concepts for describing teacher behavior, classroom situations, and pupil response have yet to be developed. It is the author's contention that this problem can best be attacked through study of electronically recorded examples of classroom interaction.

A third matter is context. Gump, Rosencranz and Biddle, and Peterson urge that problems of teacher effectiveness be considered in contexts ranging from history to teacher manipulation of classroom situations. None of these authors has proposed a method by which important contextual variables may be isolated; none has stated formal propositions relating contextual variables to problems of effectiveness. Nevertheless, their point is well made, and the investigator who ignores situations and contexts does so at his own peril.

We are, in a sense, at the same crossroads faced by leadership research two decades ago. Prior to 1940, much conflicting evidence had been gathered on the "traits" and "behaviors" of leaders. About 1940, responsible investigators began to emphasize the contextual necessities of leadership, particularly those demands that were placed upon the leader by the needs of the followers and the organization in which leadership was exerted. For two decades, then, vigorous research has taken place on the relationships between leadership and contexts, while textbooks have emphasized the inability to make general statements about leaders. Most recently, however, such authors as Stogdill (1959) and Mann (1959) have begun the synthesis of two decades of careful empirical study only to discover that there are both generalities and specificities in what makes for effective leadership.

These syntheses came from many empirical studies. Unfortunately, there are few studies to date of educational contexts. More researchers must be recruited to the study of contexts and their effects upon classroom interaction and educational outcomes.

A fourth problem mentioned in all papers is the complexity of the effectiveness problem and the lack of definitive information at present relating to selecting, training, or evaluating for effectiveness. Although only a few classes of variables are closely related to the effectiveness of teachers, little distinct evidence relates these classes to one another. In many cases, it is not known which variables are important for further study.

This suggests that selection, training, and rating programs presently in operation have little to recommend them for empirical knowledge of competence. The school of education may be forced to choose among a superfluity of potential teachers; the teacher educators may have to choose between educational alternatives in classes or skills required of their students; the principal or superintendent may have to institute a system of rating in order to satisfy his school board, parents, or teachers. Each of these groups, facing real problems, can operate as well extemporaneously as they can by study of the best contemporary research on competence.

What can the concerned teacher, administrator, or parent do to promote teacher excellence? The complexities and "basic" character of research on effectiveness have been emphasized. Is there nothing that the local school system can do to raise the standards of education in its community? Is all local research worthless, the field to be left only to the experts?

Perhaps the most important thing the citizen can do is to cooperate with others in support of research on teacher effectiveness. This can be done at several levels. First of all (through support of national and regional education associations) research can be mobilized for sophisticated attacks on the effectiveness problem. Secondly, the local school system can volunteer cooperation with problems of long-term research conducted through a nearby university and other agencies engaged in educational research. Finally, and perhaps most important, the school can adopt an active and inquiring approach to educational experimentation—which may lead to the creation of effectiveness, even if it cannot yet be measured adequately.

This last point should be emphasized. Productivity, morale, and quality are likely to depend on the amount of freedom granted, whether one is dealing with citizens in a community, workers in factories, or teachers in schools. Teachers who are constrained in rigidly defined situations or who work in a system with inflexible, hierarchical controls will not have an opportunity for development. The school that encourages experimentation with education contexts, new approaches, new facilities, new classroom situations is likely to create an excitement for education that guarantees learning—even by the teacher. How many schools are experimenting with team teaching, television, teaching machines, classroom size, variable class lengths, supplementary training programs, seminars, new curricula in mathematics, outright discussion of racial problems, foreign visitors, intramural sports, new types of library facilities, and so on?

The necessity for a serious "push" in effectiveness research must be continually recognized. The problems outlined in this book are finite and solvable, whether the aim is to select, to train, or to evaluate teacher effectiveness. Recognition of the fact that these problems are complex and beyond the scope of most previous research stimulates despair or a spirit of challenge. The latter is preferable. It is the author's position that several of the contributions presented here represent steps forward (certainly *none* of them could have been written ten years ago). The next decade should see an elaboration of these conceptual systems, the development of others, the collection of many times the present amount of hard data on issues relating to teacher effectiveness (and more particularly classroom interaction), and the beginnings of formalized propositions tying together the classes of variables laid out in this chapter.

2

HAZEL DAVIS

Evolution of Current Practices in Evaluating Teacher Competence

Methods of evaluating teacher competence currently used in public school systems have evolved from practices of many years ago. It seems clear that these evolving practices have been related to certain movements in government, in industry, and in psychological research.

The purpose of this chapter is to report briefly on the history and present status of these procedures. Although evaluation in school practice is essentially part of the supervisory process, the effort is made in this chapter to consider evaluation separately. Only cursory attention is given to supervision as a whole. Major attention is given, within the area of evaluation, to the practice of applying judgments that result in officially recorded formal reports.

The processes of making the judgments and the forms on which they are recorded were once known as "rating" and "rating forms." The current terms are usually "evaluation" and "evaluation forms," although other phrases are used. An argument might be made for identifying as "ratings" only the classifying judgments that group teachers in several different levels of competence, but this is not common practice. The text reflects changes in terminology occurring over the past fifty years by discussing "rating" in the early years and "evaluation" at present.

Supervision as Related to Teacher Evaluation

During the nineteenth century (in the first half century of evolving the position), the total responsibility of a superintendent of schools frequently was referred to as "supervision." This word eventually took on a more limited meaning, however, to identify only the responsibility for overseeing and improving the classroom work of teachers and for organizing the materials of instruction. Arguments on the functions of the supervisor began to be heard after about 1910.

It was claimed that some supervisors were serving as inspectors rather than helpers; that they laid down prescriptions regarding methods and materials, rather than providing teachers with a variety of suggestions and a wealth of materials from which selections could be made in accordance with the teacher's judgment; and that the supervisors found fault but failed to offer constructive ideas, demonstrations, or statements of principle that would help the teachers improve their work (NEA, Proceedings, 1917, pp. 187-194). The alternative following each of the above criticisms represented, in general, ideals accepted by most supervisors (who have continued their efforts to develop truly creative programs of supervision and who have continued to increase in numbers and in the scope of their services).

Supervision, in the specific sense of concern with the quality of the instructional process, has always been recognized as a responsibility of school principals. Also, administrative theory sees the supervisor who visits the school building as a staff worker subordinate to the principal when in the school. Although most American school systems do have one or more supervisors, they still are not universally employed, and thousands of teachers are supervised only by a principal. Other teachers may receive a limited degree of supervision from a superintendent of schools.

Where there is a supervising principal, he is almost always the chief evaluator of his teachers. The principal has the major responsibility in deciding whether or not a teacher is to be reemployed, to be recommended for continuing tenure, for a salary increase, or for promotion. Sometimes a supervisor is also consulted. Those who urge greater use of formal evaluation of teachers point out that some method of appraisal is already applied, by one or more supervisory officials, to every teacher in the public schools. This cannot be denied.

There are differing views, however, on which methods of evaluation are most desirable and how much they can be depended upon. Some believe that a general impression reported through an informal judgment means as much as a more analytical series of impressions recorded on a formal instrument. Other arguments suggest the value of a formal instrument that the evaluator and the teacher use only as a basis for analysis and the improvement of instruction—as compared with a formal instrument used to arrive at a classifying judgment that measures the teacher's total worth against that of other teachers.

Once teachers have met the test of a probationary period, have been accepted as qualified, and are satisfactorily carrying a regular teaching assignment, many supervisors feel they should not be classified at first-rate, second-rate, and so forth. They feel that useful evaluation does not label teachers but recognizes each as having strengths and weaknesses, and helps each to reinforce his strengths and to overcome his weaknesses.

Formal Evaluation

Formal evaluation of teaching, as practiced today, appears to have had its origin, in part, during late nineteenth-century school practice as well as in the efficiency movement of the early twentieth century.

The Efficiency Movement

Early in the 1900s, *scientific management* swept the country, with "efficiency" its catchword. Frederick Winslow Taylor's pioneering work—for the Bethlehem Steel Company and other industries in the 1890s—with his emphasis on *standardization, systematization,* and *stimulation,* laid the foundation for "the efficient age" (Kaufman, 1913). During the period from about 1905 to 1915, books were published on such topics as efficient philanthropy, efficient living, efficient democracy, and the spiritual sources of efficiency, in addition to many publications on efficiency in industry. Early in this period the president of Ohio University devoted a baccalaureate address to efficiency (Jelke, 1907). He remarked that the word "strenuous" was being supplanted by "efficient," and said that the meaning of "efficient" is "composite with the significance of the good, the strenuous, and the intelligent."

The Efficiency Society began its journal and proceedings in 1912; *Efficiency: The Magazine of Management* also began publication in 1912. The Academy of Political Science devoted its 1913 proceedings to the topic of efficient government. Taylor's own master work, *The Principles of Scientific Management* (1911), was not published until four years before his death. His theories and those of his colleagues and disciples continued to be widely influential, however, through efforts of the Taylor Society and the rising concept of industrial engineering in employment practices.

The social survey, the school survey, and the rise of measurement in education were related movements of the same period and were given greater acceptability by the obvious analogy to scientific management in industry. Community surveys of all types swept the country (Kellogg, 1909). A bibliography published in 1931 listed 2775 survey reports, 627 of which dealt chiefly with education (Harrison, 1931).

In 1912, the Commissioner of Education for New Jersey reported to the NEA Department of Superintendence his impressions of about a half-dozen efforts to determine school efficiency by means of survey investigations (NEA, Proceedings, 1912, pp. 376-381).

In 1912, also, the measurement movement was identified by the organized teaching profession in the first report to the National Council of Education (part of the NEA) of its Committee on Standards or Tests for Measuring the Efficiency of Schools or School Systems (NEA, Proceedings, 1912, pp. 560-566). George D.

Strayer was chairman of the Committee. In discussing this report, Lotus D. Coffman urged that the "widespread movement to reduce every type of productive activity to scientific management" be recognized as transitory, with its most productive outcome likely to be an improved method of defining existing standards and developing new ones (NEA, Proceedings, 1912, p. 571).

School surveys placed emphasis on testing the efficiency of teaching, most often utilizing the newly developed standard tests in school subjects, such as arithmetic, handwriting, and others. There was little interest in individual tests of individual teacher efficiency. The survey movement did succeed in keeping the idea of efficiency in the forefront of educational concern. However, the growing use of individual efficiency ratings for teachers seems to have been stimulated by the efficiency movement, by interest in educational measurements, and possibly by fear of the surveys rather than by direct use of teacher ratings by the survey teams.

Employee Ratings in Business and Industry

The effect of the efficiency movement on the individual worker in industry and business was seen in time-motion studies, the setting of work quotas, wage administration, the use of tests in selecting and assigning employees, analyses of individual output and errors, and the like. The type of evaluation commonly referred to in the literature of business as "merit rating," however, did not develop to any marked degree in private employment until after 1915

The rating sheet for sales people at Lord and Taylor, first used late in 1916, was practically without precedent. The individual responsible for it had looked about for an example to imitate, but found none. Apparently, he was unaware of the many examples to be found in public school systems. The Lord and Taylor rating form listed several traits, such as health, integrity, and industry. Shortly after this rating sheet was put into use, it was submitted to an informal group of psychologists and employment men who met periodically in New York City. They rejected the idea as unsound, but a year later reconsidered and expressed at least a measure of approval (Rayburn, 1944).

Walter Dill Scott was chairman of the Committee on Classification of Personnel in the Army in 1917 when the first rating scale for officers was introduced (Scott and Clothier, 1923). This was a man-to-man scale. The raters were directed first to construct their own scale of excellence by listing in order twenty-five or thirty officers of their own acquaintance. They were then to classify each ratee as matching, as nearly as possible, one of the men on the rater's scale. These two procedures were to be followed in regard to each of five qualities being evaluated. After World War I, Scott developed a graphic rating scale that has been the foundation for many similar ones to follow. One of these was the much-copied Probst scale (Probst, 1947).

Formal merit rating has not gained the general acceptance in industry that has often been supposed.[1] One large survey, reported by the National Industrial Conference Board (1936), showed that 363—or only 14.8 percent—of 2452 companies investigated, maintained rating systems. The bulk of these were large companies, employing 38 percent of the total number of employees.

The Bureau of National Affairs, after surveying a representative panel of personnel executives, reported (1957) that formal merit rating of at least some rank-and-file employees (other than probationers) was carried on by about 40 percent of the companies surveyed. Of the companies using merit ratings, only 29 percent of the companies having as many as 1000 employees and 39 percent of those with fewer than 1000 employees used merit ratings in connection with wage increases. Roughly one third of 40 percent is a rather small use of merit ratings in wage administration in industry. Samples of the rating forms used were reproduced. They were predominantly graphic rating scales, with space for comments.

As in public education, a running controversy has continued as to the desirability, the reliability, and the validity of merit ratings in business and industry. It seems, however, that the article entitled "Whatever Became of Merit Rating?" was premature in assuming the disappearance of the practice (American Management Association, 1958).

---

[1] It seems that school boards, especially those in large school systems, have made more widespread use of efficiency ratings than have industrial employers. See Table 1 in this chapter.

Teacher Ratings

Reports consulted in the preparation of this chapter did not include a review of the earliest devices used for rating public school teachers. A form in use in Milwaukee in 1896 consisted of a long list of unclassified traits (Adams, 1919). In the early 1900s, some of the large city systems were giving teachers numerical efficiency grades. Superintendent Cooley of Chicago, in 1907—with apparent amusement as well as exasperation—referred to the insistence of school principals on giving high marks to teachers (NEA, Proceedings, 1907, pp. 94-101). In Brooklyn, he said, the marks of over 90 percent of the teachers were over 90 percent. In Chicago, 96 percent of the teachers were marked so high as to entitle them to promotion.

E. C. Elliott, professor at the University of Wisconsin, released in 1910 the "Provisional Plan for the Measure of Merit of Teachers." This was highly influential (Boyce, 1915, pp. 78-81). The form proposed by Elliott seems to have been a genuine social invention. It was a score card so widely copied that it now appears commonplace. It included seven headings—physical efficiency, moral-native efficiency, administrative efficiency, dynamic efficiency, projected efficiency, achieved efficiency, and social efficiency. Each of several subitems under each heading was assigned a maximum value; the points totaled 1000 for the score card as a whole.

The analytic breakdown of this check list and its use of the word "efficiency" identified it with the scientific management movement. The author suggested that the plan was not intended for use by inspectors, but should be placed in the hands of teachers for their own guidance—to serve as a basis for cooperative effort between teachers and supervisors. Frequent references in articles and addresses of subsequent years indicate that the Elliott plan created much interest.

In 1920, Rugg, who had worked with Scott in developing the man-to-man scale for Army officers, reported such a device for use in rating teachers. The reference shows a tie between teacher ratings and the personnel rating movements in government and industry, but Rugg himself later questioned the value of the scale for teachers (1921).

It is beyond the scope of this chapter to review all the plans for studying teacher effectiveness proposed by researchers during the next four decades. Suffice it to say that little further identification with employee rating in business and in general government has been apparent until recent years.

The critical incident technique developed by Flanagan for use in the U.S. Air Force has been adapted to teacher evaluation procedures in some school systems (for example, Macfarlane and Furso, 1959). Several school systems within the past decade have been using industrial management consultant service to help them develop plans of teacher evaluation. One school system recently adopted an appraisal form taken directly from industrial models. This used the equal-appearing-interval approach to develop three equivalent forms of the same instrument (West Hartford Public Schools, 1961). The items were selected and weighted on the basis of the judgments of the teachers and administrators concerned, but the relative weightings of the items, derived through statistical techniques, are not known either to the evaluators or the evaluatees. Further refinements are built into the scoring process to adjust for evaluator bias toward leniency or severity and for the tendency to give higher ratings to more experienced teachers.

Surveys of the Use of Teacher Ratings

This section refers to several teacher-rating studies, including one reported in 1915, and closes with a longitudinal series of questionnaire surveys made by the NEA Research Division. Only a limited sampling of these studies is presented.

BOYCE, 1915 A landmark report on teacher rating in public school systems is the survey of the practice, compiled by A. E. Boyce, published (1915) by the National Society for the Study of Education. The magic word "efficiency" in the title identified it with the forward-looking educational thinking of the day.

The author sent inquiries to 350 cities of over 10,000 population, requesting a statement of their methods of determining and recording the efficiency of teachers. He received replies from 242. The methods listed, and the number and percent of cities reporting each, were as shown in Table 1.

TABLE 1

Methods of Determining Teacher Efficiency as Reported by Boyce (1915)

Method	*Number*	*Percent*
1. Promotional examinations	14	6
2. Schedule of qualities on which teachers are judged	133	55
3. Efficiency grades for teachers	99	41
4. Judgment of teachers not controlled	98	41

Item 4 in Table 1 presumably was exclusive of the other three. Since promotional examinations were used only in connection with some type of efficiency rating, it appears that about 60 percent of the cities reporting in 1915 were using some type of rating or formal evaluation of teachers.

Boyce reported that the number of items on which teaching efficiency was judged ranged from as few as two items to as many as eighty (p. 17). He identified four types of analysis: (1) descriptive reports dealing with specified points; (2) lists of questions to be answered by yes or no; (3) lists of items to be evaluated by a stated classification, for example, excellent, good, medium, unsatisfactory, and (4) lists of items to each of which was assigned a definite numerical value representing the maximum score that might be given (the Elliott model).

He summarized the qualities evaluated in 50 of the rating schemes. "Discipline" led all the rest, being found in 98 percent of the forms. Next in frequency were "instructional skill" and "cooperation and loyalty," each mentioned in 60 percent of the forms.

Boyce proposed a plan of rating that included a classified list of forty-five items under five headings: personal equipment, social and professional equipment, school management, and technique of teaching (pp. 44-45). The items ranged from "general appearance" to "moral influence." A separate page provided brief definitions of each term. He proposed a graphic plan of checking each item on a scale of five terms ranging from "very poor" to "excellent." Many evaluation forms in use in 1962 are similar to the "efficiency record" published by Boyce in 1915.

MONROE AND CLARK, 1924 Monroe and Clark, writing in 1924, summarized the researches of the preceding twenty years. They cited studies that had shown the lack of reliability of existing rating devices. They pointed up the existence of a halo effect from the rater's general estimate of the teacher—this influencing the estimates of particular traits.

These authors suggested that the chief value of score cards would be as a means of self-improvement when used by teachers to consider their own work. Monroe and Clark proposed, as a substitute for score cards or man-to-man ratings, a composite evaluation plan in which controlled educational measurements of pupil achievement would be given a substantial weight. Monroe and Clark included a bibliography of eighty items.

KING, 1925 A report to the National Education Association in 1925 mentioned three surveys that showed that at least three fourths of the large city school systems were using some type of efficiency ratings (NEA, Proceedings, 1925, pp. 200-215). Most detailed of these surveys was a study by Leroy A. King of the University of Pennsylvania (1925). King, like Boyce, analyzed a sample of the rating instruments then in use in large public school systems. King (p. 19) compared the factors found in the forms he analyzed with the findings of Boyce (see Table 2).

The difficulties that beset the researcher who tries to categorize elements found on rating forms are illustrated in the foregoing list, which includes the ten items of highest frequency reported by Boyce and by King. Of the fourteen categories named, only eight appear on both lists. However, King's "classroom management" and "class discipline" are both included in Boyce's "discipline." "Personality," as reported by King, may include both "manner" and "voice," reported separately by Boyce and not mentioned by King.

Words such as "excellent," "very good," and so forth; or letters such as *A*, *B*, *C*; or ranking numbers such as *1*, *2*, *3*; or per-

TABLE 2

Comparison of Items Found in Rating Instruments by Boyce and King

	Percent of rating forms	
	Boyce (1915)	*King (1925)*
Quality or factor rated	*50 forms*	*103 forms*
Technique of instruction	60	96
Personality	40	86
Classroom management	——	62
Teaching results; pupil achievement	16	58
Professional attitude	32	57
Scholarship and professional training	64	56
Class discipline	——	37
Cooperation	60	32
Daily preparation	26	31
Health or vitality	22	28
Discipline	98	——
Plan and method	56	——
Manner	30	——
Voice	28	——

centages; or descriptive phrases—any of these may be used as symbols by which judgments regarding teachers' efficiency are recorded on rating forms. King tabulated the relative frequency of these various devices (1915, p. 20); Boyce had made a similar analysis (1925, p. 44). The two distributions are shown in Table 3.

TABLE 3

Breakdown of Types of Symbols Used To Evaluate Teacher Efficiency, Reported by Boyce and King

	Percent of rating forms	
	Boyce (1915)	*King (1925)*
Symbol used in recording judgments	*50 forms*	*103 forms*
Single words or descriptive phrases	57	52
Letters	18	14
Figures (rankings)	10	23
Percentages (weighted point scales of the Elliott type)	15	11
Total	100	100

King noted also that of the 103 forms he studied, 13 percent were of the graphic type proposed by Boyce (and later introduced in industry by Scott).

REAVIS AND COOPER, 1945 Some twenty years after the King study, a comprehensive report on the evaluation of merit in city school systems was issued by Reavis and Cooper. These authors secured rating devices and accompanying instructions from 103 school systems. One of the many valuable features of this report is a discussion of terminology used by various students of the topic and the effort to produce a classification of mutually exclusive types of rating methods. They identified five. Although a given rating sheet may include two or more sections in which two or more of the five types are included, the methods appear to be distinct, as follows:

1. CHECK SCALE lists several attributes, functions, or outcomes, each of which is *evaluated separately* according to some specified series of values. Graphic scales, weighted point scales, and lists of items to be answered by yes or no all fall within this large group.
2. CHARACTERIZATION REPORT differs from the check scale in that only the *total* merit or work of the teacher is evaluated, according to a specified scale of values. Instead of a series of factors to be rated, a single rating is given to the teacher.
3. GUIDED-COMMENT REPORT lists a series of topics or questions on each of which the rater is required to write a comment relative to the teacher.
4. DESCRIPTIVE REPORT requires a descriptive statement (it may be headed "comments" or "general evaluation") but does not specify topics or emphases and leaves the structure of the statement entirely to the discretion of the rater.
5. RANKING REPORT requires the rater to list the teachers of a given school unit in order of excellence, placing the best teacher at the top and continuing until all teachers are included.

Although these types can be distinguished from each other, the systems studied by Reavis and Cooper were not classified in a 100 percent distribution because some systems used two or more different rating forms, and many of the forms included two or more of the five methods. They reported the following frequency distribution with percentages based on the total of 103 cities: (1) check scale, 75 percent; (2) guided comment, 31 percent; (3) characterization report, 24 percent; (4) descriptive report, 6 percent; (5) ranking report, 1 percent (p. 18).

These investigators recorded the number of scale values used in check scales and found (like King) that a five-point scale was a median of sorts, representing as it did 42 percent of the check scales analyzed (p. 35).

The symbols for recording the values in check scales were noted —31 percent used numbers; 82 percent used words or general phrases; 9 percent used definitive or descriptive phrases (p. 37). This distribution shows that 22 percent of the forms used more than one of the three types of symbols.

Reavis and Cooper, with other investigators (such as Barr and Emans, 1930), analyzed specific items in the rating forms. Their categories were so dissimilar from Boyce's and King's as to make comparisons impossible. Table 4 gives a classification of items according to seven broad groups, its percentages based not on the number of scales analyzed, but on the total of 1538 different items noted in eighty-five scales (pp. 31-33).

TABLE 4

Items Found in Rating Instruments by Reavis and Cooper

Quality or factor rated	*Percent of items*
Social relations	21
Instructional skills	19
Personal characteristics	15
Noninstructional school service	14
Professional qualifications	11
Habits of work	10
Pupil results	10
Total	100

Reavis and Cooper devoted more than ten pages of their report to a review of the value of rating forms of the types they found in common use in school systems (pp. 81-93). Certain weaknesses were apparent—lack of definition of items, ambiguous terms, and items that linked two independent elements for a single judgment. Of 1538 items recorded in eighty-five scales, 1282 items appeared only once. It was the judgment of the investigators that the scales neither followed accurately nor sampled adequately a pattern of general ability.

The discussion included brief summaries of a few of the studies, for example, those by Rugg (1922), and Barr (1929), plus later reports. These show lack of agreement among raters, plus the failure of supervisory ratings to correlate with measures of pupil achievement. On the basis of this review of the lack of usefulness of rating forms, Reavis and Cooper concluded (p. 87):

> Ratings appear to be invalid, then, as comprehensive measures of either general or specific teaching ability. There still remains for them, however, the sphere of personal opinion. . . . It is apparent that ratings are valid measures of the rater's opinion of a teacher.

These investigators advance the theory that teacher ratings have value chiefly as measurements of a teacher's prestige and that other devices should be used to measure other aspects of teaching success. Among these, they mentioned observational records of classroom technique as a promising new development, although still dependent to a large degree on subjective reactions.

NEA RESEARCH DIVISION, 1923-1962 The Research Division of the National Education Association has sent a number of questionnaire inquiries (including items on teacher ratings) to urban school superintendents. The first was reported in 1923. The most recently published was for the period 1955-1956.

Table 5 summarizes the replies to a question on whether or not classroom teachers are rated. Prior to 1950-1951, the wording from study to study was not identical although the word "rated" or "rating" is central to each question. It is impossible to know whether or not varying definitions of the word "rating" might have biased the findings so as to induce a higher response in the earlier years than in subsequent years. It is believed, however, that the figures may be reported as a continual series. The precise wording of each question is quoted in Table 5.

From 1923 to 1956 the percent of districts in which all teachers were rated declined from 55 percent to 39 percent. The percentages are weighted averages, based on replies from urban school districts of various sizes (2500 population and up). The weighting is based, not on number of teachers, but on number of school systems; substantial samples were reported in each population group in each survey.

The reduction in the proportion using ratings was less marked in small systems than in large ones. However, in all years large cities reported the use of ratings by higher percentages than was true for small systems. The reporting districts with population of 100,000 and over showed 82 percent using ratings in 1923; this figure was down to 39 percent in 1955-1956. The smallest systems, 2500 to 5000 in population, showed a decline from 51 to 35 percent.

Preliminary data from a 1961-1962 survey show a marked *increase*, in the six years following 1955-1956, in the percent of urban school districts in every population group reporting the practice of rating all teachers. The percentage of districts, over 100,000 in population, in which all teachers were rated was 39 percent in 1955-1956. This rose to 58 percent in 1961-1962. The increase was also from 39 percent to 58 percent in all urban districts.

Not shown in Table 5 is the great difference between large and small systems in the percentage that rate *some* teachers (probationers or teachers believed to be failing) and the percentage that rate *no* teachers. In 1961-1962, of 145 districts reporting from large communities 100,000-499,999 in population, 56 percent rated *all* teachers, 37 percent *some* teachers, and only 6 percent *no* teachers. Of 186 small districts 5000-9999 in population, the corresponding figures were 55 percent, 5 percent, and 32 percent.

In several different years the survey has included a question on the uses made of the ratings. In 1955-1956 the highest frequencies were for the items listed in Table 6 (NEA, Research Division, 1956, pp. 28-29).

Use of ratings as a supervisory aid would suggest that the teachers should be given copies of their ratings as a basis for study and efforts toward improvement. Thirty-five percent of the districts in which teachers were rated reported in 1961-1962 that teachers were not given their ratings, and 9 percent left the question blank. This 44 percent may include some places in which the

TABLE 5. Trends in Use of Teacher Ratings in Urban School Systems

Year when study was made and wording of question	Percentage of affirmative replies, by population of school district							
	500,000 and over	100,000–499,999	100,000+ (*cols.* 2–3)	30,000–99,999	10,000–29,999	5000–9999	2500–4999	*Weighted total**
1922–1923: Are your teachers rated as to quality of service rendered?	†	†	82	72	57	54	51	55
1927–1928: Have you a plan under which all teachers are definitely rated as to quality of service?	†	†	71	64	37	42	38	41
1930–1931: Are all teachers given a definite rating periodically as to quality of service rendered?	†	†	71	59	46	43	32	41
1940–1941: Are all teachers regularly given definite ratings on the quality of service rendered?	†	†	67	50	41	36	35	39
1950–1951, 1955–1956, and 1961–1962: What members of the professional staff are rated regularly on quality of service? (Percents represent replies indicating that *all* classroom teachers are rated.)								
1950–1951	68	53	55	43	46	41	39	42
1955–1956	59	36	39	41	42	37	35	39
1961–1962	74	56	58	63	61	55	51	58

* Weighted percents differ slightly from those originally published.

† Districts 100,000 and over in population were shown as a single group in 1940–1941 and earlier years.

SOURCES: 1922–1923—*NEA Research Bulletin* 1: 52; 941 districts reported. 1927–1928—*NEA Reasearch Bulletin* 6: 239; 1532 districts reported. 1930–1931—*NEA Reseach Bulletin* 10: 65; 1465 districts reported. 1940–1941—*NEA Research Bulletin* 20: 96; 1737 districts reported. 1950–1951—*NEA Research Bulletin* 30: 49; 1597 districts reported. 1955–1956—NEA Research Division Special Memo, June 1956, p. 28; 1973 districts reported. 1961–1962—Preliminary data from questionnaire issued November 1961.

teacher sees and signs the rating, but does not receive a copy; the stress of such a situation, however, would make it difficult for a teacher to remember the details of the rating he sees.

TABLE 6

Uses Made of Teacher Ratings in 1955–1956

Uses reported	*Percentage of 1064 districts using ratings*
As a supervisory aid	77
In deciding on reappointment of teachers not on tenure	71
In recommending probationary teachers for permanent tenure	56
In selecting teachers for promotion	36
In paying regular increments on the salary schedule	16
In selecting teachers to receive supermaximum salaries	7

In the 1955-1956 study, the question was asked, "Who does the rating?" In 37 percent of the districts the principal was the only rater (NEA, Research Division, 1956, pp. 28-29). The larger the city, the more likely this was to be true. Fifty percent of the very largest districts (a half million and over in population) gave entire responsibility for rating to the principal (as compared with 35 percent in the small districts below 5000 in population).

Joint ratings by principal and supervisor were reported by 27 percent of all districts, and separate ratings by principal and supervisor in 17 percent. Various other groupings of officials, usually with the principal as one member, accounted for the remaining 19 percent. The principal is clearly the primary rating officer.

Evaluation Forms Used in 1961–1962

The Research Division of the National Education Association in November 1961 sent a questionnaire on personnel practices to a large number of school superintendents. The inquiry included a request for copies of forms used in the evaluation of teachers. In the nearly forty years since the Research Division issued its first inquiry on this, the number of school systems in cities 100,000 and over in population has increased from 67 to 189. Rating forms had been received from 81 of these 189 systems at the time this chapter was written.

This section presents some comments on these documents—comments based on an impressionistic rather than a statistical analysis. The first reaction was bewilderment at the infinite variety, and the second, a greater awareness of the difficulties faced by earlier researchers who have tried to present an ordered summary of such forms. A count was made of only two aspects of the forms: (1) the methods of evaluation used, and (2) in the case of check scales, the number of intervals in the scale of possible judgments set by the forms.

Many of the eighty-one districts sent more than one form. There are forms for a preliminary evaluation and others for a final evaluation. One sent separate forms for elementary school and secondary school teachers. A few sent different forms for probationary and permanent teachers; several sent a form only for probationers. Some sent observation records and discussion outlines for evaluation conferences. In the count reported below, only one form was used from each of eighty-one school districts—the final form, where there was a sequence; the form for permanent teachers, if there was also a separate one for probationary teachers; or the form for probationers, if that was the only one. Forms designed for use only in self-evaluation were not included.

The methods of evaluation found in the forms were classified (with slightly modified wording) on the basis of those used by Cooper and Reavis. Four types of evaluation were noted: (1) multiple-factor check scales, (2) general-factor check scales, (3) structured comments, and (4) nonstructured comments.

The distribution of the eighty-one rating forms to show the use of the four types of evaluation is shown in Table 7.

Eighty percent of the forms included multiple-factor check scales, and 40 percent included no other method. The general-factor check scale used alone was extremely rare and, even in combination with other methods, was the one least used. Structured and nonstructured comments were seldom used alone but were found as a supplement to other methods.

An exact count was not made of the evaluation forms in which

TABLE 7

Methods of Evaluation Used in Eighty-One Evaluation Forms, 1961-1962

Method of evaluation	*All forms examined*	*Forms using only one method of evaluation*	*Forms using two or more methods of evaluation*
Number of evaluation forms	81	45	36
Percentage of total	100	56	44
Multiple-factor check scales:			
Without a general evaluation	57%*	25%*	32%*
With a general evaluation	23	15	9
Total multiple-factor check scales	80%	40%	40%
General-factor check scales	14	4	10
Structured comments	32	8	24
Nonstructured comments	23	4	19

* All percentages are based on N = 81.

SOURCE: Evaluation forms received in December and January of the school year 1961-1962 from urban school districts 100,000 and over in population.

a single classifying judgment was recorded comparing teachers with one another on relative levels of efficiency. However, the proportion would be not more than 37 percent (the multiple-item check scales with a general evaluation, plus the general-factor scales, allowing for no duplication). The process of evaluation without a comparative classification is in line with a principle stated in the introduction to the Cincinnati evaluation form (Cincinnati Public Schools, 1959):

> This form may be used by the teacher for self-appraisal or for the principal's appraisal of the teacher. It is understood that all teachers using the form will check some characteristics as their weakest, and that this is simply an indication that improvement is desired, not that performance is unsatisfactory. Checking in this fashion should not cause concern regarding the accuracy of the checks, nor the interpretation others may place upon them, for self-appraisal seeks only to indicate what a teacher feels about his own performance, and to stimulate the growth which comes from a systematic review of desirable teaching practice. The code is specifically designed to *prevent comparison of one teacher with another.*

In multiple-factor check scales, a series of items is listed, and for each item a space is provided for recording a judgment, according to a scale stated on the form. The items to be evaluated may be stated as single words, such as "judgment" or "discipline"; they may be phrases describing characteristics or conditions, such as "ability to arouse pupil interest," or "evidence of careful planning"; or they may be descriptive sentences, such as "helps children acquire good study-and-work habits." The items are usually (but not always) classified under several major headings—which themselves may or may not be the subject of evaluation. A substantial minority of the multiple-factor check scales include, as one item, a general evaluation of the teacher on the same scale as for specific items.

The items to be evaluated usually are substantives or sentences; the value judgments in the scale usually are adjectives. Nearly half the check-scale forms use a five-interval scale or code, usually one of the many variations of the familiar sequence, "excellent, very good, good, fair, poor." Slightly less than one fourth use a four-interval scale and nearly one-fifth, a three-interval scale. The three-interval scale is usually "satisfactory, doubtful, unsatisfactory," or some variation thereof. The middle interval in the three-way scale is "needs help" in a group of forms that appear to have been influenced by the plan used in Cincinnati.

Some scales provide an extra code, beyond the scale, for "insufficient opportunity to evaluate" or some similar phrase. This adds a helpful note of realism as some of the scales include items that appear to defy evaluation. These are such undefined traits as "naturalness," "attitudes," and "scholarship"; also double-headed items such as "promptness and accuracy," and "control and influence."

The physical recording of the judgment in a multiple-unit check scale usually is done by checking one of a series of spaces; a few require the writing in of a number. Where numbers are used they represent, generally, a stated verbal equivalent. Only one of the forms includes a weighted percentage scale, as in Elliott's score card. King noted 13 percent of scales of this type; Reavis and Cooper, 11 percent. Only a very few provided a graphic scale designed for the preparation of profiles (as in Boyce's, Scott's, and Probst's plans).

In the general-factor check scale, only the total worth of the teacher is evaluated in a single judgment. About one third of the multiple-factor check scales include a general over-all evaluation that would be a general-factor check scale if it stood alone. One or two forms include both a general evaluation of a multiple-factor check scale and also a general-factor scale using a different code.

A number of the forms call for self-evaluation by the teacher, usually to precede an evaluation by the principal. The teacher enters a scale symbol in one column, and the principal records his judgment in a parallel column. An interesting feature in a few of these forms is the provision of two scales—one for the teacher and a different one for the principal. Whereas the principal chooses among the usual sequences of ranking values, the teacher checks each item according to a code similar to the following: (1) one of my stronger characteristics, (2) neither one of my stronger nor one of my weaker characteristics, and (3) one of my weaker characteristics. Many forms provide space for the signature of the teacher as well as the principal. The teacher's signature is usually accompanied by a disclaimer of approval or disapproval of the evaluation.

The above still does not bring us to the heart of the matter: the kind of teacher (or teaching) being held up as a model by the evaluation form. A common criticism of evaluation practices is that they fail to define the scope of the teacher's total task as a basis on which the teacher may orient his efforts and judge his success. This would require the preparation of a detailed job description of each teacher's annual assignment because each teacher's task is in some ways unique each year and presents different requirements. If the use of an evaluation form is assumed under such a plan, the form should be individually designed and changed, so as to cover the whole duty of the teacher.

It may be that, in school systems not using evaluation forms, an ideal such as the above—but without formal writing of job descriptions and appraisal forms—governs the evaluation process. Inspection of the forms did not reveal any that seemed related either to a concept of job evaluation or to a formal application of role theory to the work of the teacher.

A majority of the check scales classify the items under headings designed to cover broad areas of a teacher's responsibility. The forms that call for structured comments usually specify items similar to the major headings in the check scales. Short of actual duplication of forms, of course, no two outlines are alike, and the number of times under each heading varies indefinitely. A few forms have only two main headings, such as "personal qualities" and "professional qualities," while others may go to six or seven. Usually, the headings cover such concepts as personal qualities and relationships, professional background and relationships, instructional skills and classroom management, and community relationships. Instructional functions are frequently analyzed under more than one heading.

A few forms provide space for a description or characterization of the year's assignment. A few instruct the evaluator to describe unusual conditions and to hold the teaching situation in mind in making the judgments. In general, however, such conditioning factors are ignored in the forms.

About a half dozen of the forms were accompanied by booklets (several of which were committee reports) setting forth in some detail the objectives and philosophy of the evaluation process. These documents outline the specific responsibilities of the evaluators, the timing, and certain procedural steps. A much larger number of the forms include definitions and a few details on procedures.

A long-standing issue among research workers has been the question as to whether evaluation should concern itself with *the teacher* or *the teaching*. In practice, this issue is clearly resolved: *the teacher* is evaluated. All of these school systems seem to place primary emphasis in evaluation on the teacher as a mentor and counselor for youth. Most of them also include concern for the teacher as a person, as a member of a school staff, as a member of a profession, and many also evaluate him as a member of the community.

Evaluation Other than Through Formal Instruments

The emphasis on formal instruments in the preceding sections of this report may be overdone as far as the total program of teacher evaluation in public school systems is concerned. Many school systems—24 percent of all those reporting in 1961-1962—use *no* formal ratings. Even in systems having merit-pay plans, where the highest salaries are reserved for teachers judged to be superior, a substantial proportion use only informal judgments for selecting the superior teacher.[2]

In some school systems, teachers and principals alike regard the formal evaluation only as an administrative ritual required by the downtown office. Many classroom teachers go through the school year without once having their teaching observed either by a principal or a supervisor. Where schools are very large, a classroom teacher may have only rare personal contacts with the principal. In such a situation, it is meaningless for a principal to say at which of five levels of performance the teacher "uses class time economically" or "shows ingenuity in finding and using appropriate teaching techniques."

Where the evaluation is only a formality, it is of little value to anyone, and is likely to do harm through attaching negative labels instead of giving constructive advice where needed. A sense of real injustice develops among teachers when ratings arrived at in a perfunctory manner become the basis for salary increases.

It was this sense of injustice that led the members of the National Education Association, as early as 1915, to adopt a resolution in opposition to "those 'ratings' and records which unnecessarily disturb the teacher's peace and make the rendering of the best service impossible" (NEA, Proceedings, 1915, p. 29). Again, in 1961, as in other recent years, the National Education Association in its resolutions recognized that "it is a major responsibility of the teaching profession, as of other professions, to evaluate the quality of its services." The resolution also expressed opposition to the use of subjective methods of evaluation for the purpose of setting salaries, saying specifically: "Plans which require such subjective judgments (commonly known as merit ratings) should be avoided" (NEA, Proceedings, 1961, pp. 189-193).

---

[2] Of 194 school systems reporting methods of selecting teachers to receive supermaximum salaries on a basis of superior service, 22 percent used informal judgments only, and 52 percent in addition used informal judgments as one factor in the selection. Only 14 percent reported exclusive dependence upon a formal rating plan.

Evaluation, however, is not always a superficial process. In some school systems, evaluation as practiced today is making a positive contribution to better teaching. One publication on evaluating teacher effectiveness (Ellena, Stevenson, and Webb, 1961) speaks of "the need for more precise definition, observation, and analysis of the network of relationships among teacher behaviors, instructional goals, and pupil characteristics." School systems that have made a serious effort to develop adequate programs of evaluation find themselves restating instructional goals and considering not only pupil characteristics but other factors that help to determine appropriate teacher behaviors. Some of the school systems that have given the most study to the problem and have recognized the complexities of the process are among those least willing to reduce teacher evaluation to the basis of a general-factor classifying judgment.

Even in the absence of central-office concern, evaluation is sometimes given in which the supervisory evaluator informs himself on what the teacher is being called upon to do and how he is doing it and offers constructive suggestions to the teacher for doing it better. Extensive planning and administrative action are necessary, however, to make evaluation of this type an objective for all teachers in a school system.

A comprehensive personnel program, highly selective in the appointment both of new teachers and of new principals and supervisors, is essential. Programs of continued study on aims and methods of evaluation by principals and supervisors are found in only a few school systems today. Such study would seem to be a *sine qua non* for anything approaching a uniformly high level of evaluation in a school system.

Administrative staffing ratios need to be re-examined. Principals are often accused of devoting their time to "administrivia" instead of instructional leadership. However, those who have stood in the principal's shoes know how demanding the administrative processes are. Many details cannot be neglected without endangering educational values. These cannot be delegated unless someone is available to whom they can be delegated, a person with the ability and the time to handle them. Only 5 percent of the instructional staff of the public schools are principals and supervisors (NEA, Research Division, 1961a, p. 10). Creative evaluation may have to wait for further study of the work load of administrators and supervisors.

Evaluation in any comprehensive sense means consideration of the teacher's record and work as a whole. Cumulative, composite records of many types of evidence are needed as a basis for the kind of evaluation that identifies the teacher's unique abilities and strengths, the areas in which the teacher may need help, and the activities that will lead to continued growth. Reavis and Cooper conclude their report on evaluation of teachers by recommending better cumulative records and suggesting several types of information on teachers that should be on file (1945, pp. 102-125).

In school systems that take evaluation seriously, information on factors that may be related to successful teaching (commonly considered at the time of appointment) is recorded and kept up to date. Included are such items as academic and professional study, scores on written examinations, experience in teaching and related work, research and writing, educative travel, in-service education, participation in conferences and professional organizations, and the like. Extracurricular activities, committee work, and other leadership assignments in school or community are noted. Records on health examinations, absences due to illness, and other indexes of health and vigor are kept on file.

Although there is little agreement on the best methods of relating records of pupil growth to teaching proficiency, certainly any apparent correlation between pupil achievement and teachers' work might be made part of such records.

Records of supervisory observations and conferences are preserved. Several school systems, working in cooperation with university researchers, are experimenting with the use of devices for objectively recording observations of teacher-pupil interaction in classroom situations (for example, Provo City Schools, 1961; Weber

School District, 1961). These, and less formal supervisory records, such as anecdotal records, should be retained through to the completion of each current evaluation cycle. No one item should stand alone—all should be considered—in using evaluation as a means toward improvement.

Growing participation in evaluation by classroom teachers themselves seems indicated. Members of a profession traditionally regard themselves as responsible for their own actions. Teachers, no less then other professionals, have a part to play in developing policies that govern evaluation. The New Horizons project of the National Education Association claims for the teaching profession the right to and the responsibility for developing and maintaining standards for ethical and competent practice (Lindsey, 1961, pp. 205-230). This responsibility extends from the local to the national level.

Cooperative study by classroom teachers and administrators may be expected to increase the practice (noted in the sample of 1961-1962 evaluation forms) of diversity in evaluation standards within a single school system. The areas for specific evaluation, as well as the standards, might differ for probationary and permanent teachers. A few school systems are already doing this. Differences between elementary school and secondary school requirements might well be identified. Evaluative standards for administrators would seem to be a prerequisite to sound teacher evaluation. Other instructional groups—for example, counselors and librarians—are recognized in some school systems as requiring their own evaluation standards. Such differentiation will doubtlessly increase.

Growing concern with the need for creative evaluation and its concomitant, improved teaching, will certainly mean increased concern on the part of teachers with outcomes from the current research on teaching effectiveness, of which this volume gives significant examples. The record of the past shows that there has been some carry-over in earlier years from the research in teacher effectiveness to the practices in local school systems. As present studies reach completion, and as others develop, they will offer further suggestions as to types of evidence that should be collected and considered in the practical processes of teacher evaluation.

3

DAVID G. RYANS

Research on Teacher Behavior in the Context of the Teacher Characteristics Study

This chapter will consider some of the problems involved in researching and theorizing about teacher behavior; in particular, it will describe one set of researches that together comprised the teacher characteristics study. A complete report of that study—appearing originally in a volume published by the American Council on Education entitled *Characteristics of Teachers* (Ryans, 1960)—dealt with relationships between estimates of teacher-behavior patterns observed in the classroom, inventory of *estimated* teacher characteristics, background and environmental variables, and observed pupil behaviors. The original report ran to some 400 pages (almost a third of which presented tabular data referring to measured attributes of teachers). The study involved some 6000 teachers in 1700 schools and approximately 450 school systems. Obviously it is impossible to summarize here the findings of such an extensive research undertaking. Instead, we will review the major objectives and approaches of the study and suggest implications for an understanding of teacher competence.

More specifically, in this chapter will be discussed some of the techniques employed in researching teacher behavior as well as certain findings. Implications and limitations of such research for the evaluation of teacher competence also will be noted.

An Overview of the Teacher Characteristics Study

The teacher characteristics study started from scratch, so to speak. It made relatively few assumptions about the roles of teachers. Instead, it followed a design that dictated going into the classroom to observe what transpires there when teachers and pupils react and interact in the learning environment. It proceeded to attempt systemization of the observation data thus collected, to relate those observation data to other kinds of information about teachers, to discern typical patterns of teacher characteristics in relation to various conditions of teacher status. Further, an effort was made to determine kinds of information that offered clues permitting the distinction between highly assessed and lowly assessed teachers and to investigate the interactions and interrelationships among pupil behaviors and teacher behaviors.

More specifically:

1. Prior to collecting information basic to the description of teacher behavior, there was a two-pronged need: to develop instruments for recording assessments of teacher activities in the classroom and to perfect observation procedures. Assessments were to be made by trained observers. The staff, therefore, pursued for a long time the refinement of observation methods to increase the reliability and validity of the assessments obtained. The first phase of this research and development activity was directed at the determination of (a) the kinds of behaviors that could be observed in the classroom, and (b) the particular behaviors upon which attention should be focused in our study. Employing the assessment instruments and the observation procedures that were developed, large numbers of teachers (together with their classes) were observed in an effort to obtain basic criterion data relative to certain kinds of teacher classroom behavior.

2. Having conducted extensive observations and having developed analytical, factual assessments of certain classroom behaviors of teachers, the study sought to determine the major patterns of observed teacher behavior that would emerge through factor analysis of the intercorrelations of assessments of the various observed teacher dimensions. This

determination of criterion variables relative to teacher behavior was done separately at the elementary and secondary levels of teaching.

3. Concurrently with the observation and analysis of teacher behavior, the study attempted to develop questionnaires for tapping other teacher characteristics that could not be efficiently assessed by direct observation in the teacher's classroom. These were direct-inquiry, self-appraisal questionnaires developed to obtain estimates of:

a. Teachers' attitudes (toward pupils, toward parents, toward colleagues, toward administrators);
b. Teachers' educational viewpoints (academic-centered school programs versus school programs stressing other objectives, rigid school program versus flexible school program involving pupil and parent participation, teacher-directed learning in traditional subject-matter fields versus learning directed by pupil interests and abilities—these referring, in general, to what we sometimes call traditional versus permissive viewpoints);
c. The verbal intelligence or verbal understanding of teachers;
d. The emotional adjustment of teachers.
(A "validity of response" scale also was developed, primarily as a control device to estimate the extent to which a respondent was tending to give "valid," or accurate, self-assessments as contrasted with "socially acceptable" responses.)

4. Another major interest of the study, and one that consumed a great deal of staff time and energy, involved the determination of correlates, or concomitants, of observed teacher classroom-behavior patterns. This activity involved:

a. Developing paper-and-pencil instruments consisting of pools of items hypothesized to be related to teacher classroom behavior and other personal characteristics of teachers;
b. Deriving empirically validated scoring keys for such materials through a series of response analyses, employing the direct-observation assessments and direct-inquiry devices for criterion purposes. This identification of correlates was undertaken separately for a number of teacher groups (for example, teachers of grades one-two, of grades three-four, of grades five-six, of English, of mathematics, etc.) and proceeded through a number of iterations. It culminated in a determination of the validity of the resultant "correlates keys" by means of cross-validation, concurrent validation, and presageful validation under different conditions.

5. With the data made available through systematic observation of teachers as well as administration of the paper-and-pencil test and inventories, numerous comparisons were made between the teacher groups and teachers classified with respect to various aspects of status. That is, comparisons were made (with respect to the several sets of teacher characteristics upon which the study focused its attention) for groups of

teachers of different ages, for teachers with varying amounts of teaching experience, for male versus female teachers, for teachers of different grade levels and subject matters, for teachers from different kinds of schools and communities, and so forth.

Several selected aspects of the teacher characteristics study will be discussed in somewhat greater detail.

The Assessment of Teacher Behavior and Teacher Characteristics

One very basic problem for the study concerned measures to be employed in assessment of teacher behavior and teacher characteristics. Evidence of the reliability and general efficacy of the estimation procedures followed was necessary or else the basic criterion data would be useless and any conclusions without foundation. (It is well to note that the same considerations are appropriate to programs of teacher evaluation conducted by school systems.)

The Assessment of Teacher Behavior

In approaching the assessment problem, it should be kept in mind that the primary concern of the teacher characteristics study was the "personal" and "social" behaviors of teachers as those behaviors related to classroom situations.

Many specific teacher acts may be noted in a classroom, even during a very limited period of time. Therefore, a basic assumption was made to the effect that specific teacher behaviors are not completely independent of one another. Instead, it was assumed that these possess some generality and may, therefore, be classified into a fairly limited number of classes or groups, each group being made up of related behaviors possessing a common substratum identifying that group or class. Thus, any one of a number of specific acts could be thought of as belonging to one such class of behaviors that might be called "hypercritical." Various classes or sets of behaviors that might be identified as "hypercritical," "fault-finding," "cross," "sarcastic," or "scolds a great deal," might, in turn, all be considered together as belonging to a still larger class of behaviors that could, for convenience, be labeled "harshness."

The study thus began with the assumption that sets of behaviors could be organized into hierarchies of increasing generality.

At the start it was necessary to specify, in some way, the primary set of teacher traits (or characteristics or behaviors) with which the study would be concerned. Such a list might be compiled from a series of exploratory, open-ended classroom observations; it might be assembled from reports of previous investigations and deliberations; or some such selective methodology as the "critical incidents technique" might be employed.

Actually, the teacher characteristics study used all three. It leaned particularly heavily upon the critical-incidents procedure. This approach starts with (a) the designation of teachers who are, in the opinion of individuals presumed to be in position to make a judgment, "outstandingly superior" and/or "outstandingly poor," and then requires (b) the documentation of such judgments with descriptions of actual observed behaviors believed to have contributed to the judgment of superiority or inferiority of the designated teachers. (Here one is not judging teachers to rate them or compare them with some standard or evaluative criterion; the teachers are being judged simply to provide a context from which may be derived, by content analysis of reported behaviors, certain class concepts that may help to describe teacher-behavior domains.) The critical-incidents technique is not without limitations, but it was judged to be a useful approach when gathering necessary preliminary information.

Parenthetically, it is of interest to consider the kinds of behaviors people remember about teachers and to raise the question of the relative importance of such remembered characteristics with respect to behaviors normally assumed to characterize teaching. Most teaching incidents reported (of those reviewed by the teacher characteristics study) involved personal or social teacher behaviors, even though directions had given the judges complete freedom in naming critical incidents. This was *not* exclusively the case, for some incidents did have to do with the teaching of subject-matter content or with strictly intellectual traits of the teacher. A preponderance of the reported incidents, however, related to personal or social behaviors. The question, then, might be raised: are personal or social characteristics more important than a teacher's scholarliness, the teaching procedures followed, unique demonstrations, or the content taught? One well may doubt they are more important, but they may be equally important. We question why people more often do not mention incidents involving the teaching-learning process per se. An answer may be that these behaviors are taken for granted because all teachers are presumed to carry out such kinds of activities. Perhaps they are so intimately bound to the professional stereotype that teachers are thought of as being fairly homogeneous with respect to such behaviors. It also is possible that since subject-matter presentation and direction of learning involve more "technical" considerations, most people are unfamiliar with relevant criteria by which to judge the teacher and simply forget or disregard incidents in these behavioral domains. More likely, all of us seek personal reinforcement, and it is in the area of the personal or social characteristics of teachers and other persons that we best recall events.

In any event, the critical-incidents approach was a major source of information to the study as it attempted to develop an observation schedule and observing procedure. After extensive inventorying and analysis of reported critical incidents, after study of previous classroom research and "personality" research, and after a good deal of trial and error involving classroom observation and assessment, the study staff selected a limited number of "important" teacher-behavior *dimensions* in the personal or social domain. The next step was to develop an observation procedure and useable assessment blank for use with the selected dimensions.

It was necessary to devise (1) an observation-and-assessment record form, and (2) the even more important "glossary" describing, in operational terms, the behaviors to be assessed. The resulting blank was known as the classroom observation record. Specific teacher acts exemplifying the poles of each behavior dimension were described in the glossary. In its final form (after a number of revisions) the classroom observation record mentioned four dimensions of pupil classroom behavior and eighteen dimensions of teacher behavior.

Trained observers, employing this observation aid and assessment record, assigned each teacher who was observed a value on each dimension—the values extending from one to seven. An assessment of one (at the extreme left pole) on the "harsh-kindly" dimension indicated an inference based on observation of a preponderance of "harsh" behaviors; a seven (at extreme of the right pole) indicated an inference based on the observation of many "kindly" teacher behaviors.

The success of direct observation and assessment for the identification and estimation of behavior characteristics depends to a great extent upon the ability of the observer to perceive accurately, the degree to which the procedure employed may be objectified and standardized, and the extent to which significant aspects of a situation or behavior may be identified and sampled.

Much of the ambiguity of data based upon direct observation and assessment appears to arise from the lack of common understanding and operational agreement by different observers about the feature of the behavior, product, or object under observation and the absence of a common procedure followed by different observers in observation and assessment. In view of the varying backgrounds of different persons (which predispose them to perceive the same overt behavior in different ways) and considering the semantic difficulties often interfering with agreement on the operational or behavioral meaning of a "trait name," substantial agreement among untrained observers seldom can be expected.

Therefore, observers must be carefully selected and thoroughly trained. Operational definitions of the behaviors to be observed must be clear. Furthermore, once an observer has undergone a program of training, it cannot be assumed that his observational procedures will remain uniform over a period of time without occasional checkups and retraining. Experience has shown it is necessary to reinstitute training at regular intervals to insure that there have been no shifts in the agreed-upon definition and perception of behaviors being observed.

Observer training involves, first, the provision of information (and its understanding) with respect to the behaviors to be observed and assessed; and second, practice in observing, followed by discussion and comparison of procedures and assessments of a particular observer with those of other observers. During the training period it is necessary to repeat this procedure a number of times and to continue the training (much like the training of an examiner in the use of the Binet or Wechsler scales) until approximate agreement among observers is reached. The first stage of training consists primarily of study and discussion of the operational definitions of behaviors under consideration. Practice in observing ("practice" defined as a series of observation, assessment, and comparison and discussion sessions) is directed toward the development of ability to perceive and to record perceptions accurately in a specific situation. It has as its goal *consistency* by any one observer in assessing similar behaviors from one time to another and by different observers viewing and assessing the same behavior at the same time.

The observer training procedure followed by the teacher characteristics study will not be discussed in further detail. It is, however, most important to note that only with training of observers can one expect to obtain meaningful assessments of teacher behavior. It is the only proper way one can approach teacher assessment for either research purposes or for preservice and in-service teacher evaluation.

Although over a year and a half was spent in developing the classroom observation record and glossary and the observation procedures employed, the staff felt that this cautious approach paid dividends. The study was able to report quite substantial intercorrelations between observers on different characteristics and on the teacher-classroom-behavior patterns that subsequently emerged. Reliability estimates were made of the assessments of the several dimensions of observed teacher behavior (for example, "harsh-kindly," "aloof-responsive," "stereotyped-original," "evading-responsible") based on correlations between the assessments by a first and second observer of the same teachers. These, generally, were between .50 and .60: for elementary teachers the reliability coefficients for the different specific dimensions of behavior clustered around .55, for secondary school teachers, around .60.

As might be expected the reliabilities of assessments increased as we went from the separate dimensions of teacher classroom behavior to composite patterns derived by factor analysis. Reliabilities of the indexes (scores) for the patterns of teacher behavior (based on correlations between indexes yielded by assessments of different observers who independently assessed the same teachers) ranged from .70 to .80. To further assure reliability, a minimum of two independent sets of assessments was obtained for each teacher studied. The separate ratings subsequently were weighted equally and combined to form composite assessments for each teacher on each of the bipolar dimensions considered.

In respect to classroom observation methodology, it may be noted that the reliability and validity of assessments made from direct observations generally were enhanced by: (1) attention to a limited number of relevant behavior dimensions for observation and assessment (the study probably approached the practicable maximum with some eighteen or twenty dimensions, and the assessment process probably could have been further improved by reducing the number of dimensions to perhaps eight or ten); (2) the provision of specific and unequivocal operational definitions of the characteristics to be assessed; (3) insuring that the observer was well acquainted with the behaviors to be assessed and with situations in which the behaviors frequently are manifest (acquaintance resulting from extended "training" involving use of the assessment instrument and procedure in actual classroom situations); (4) focusing of observer attention on the specified behaviors or characteristics to be assessed and carefully avoiding the influence of general impressions, unusual or dramatic behaviors, or inferences about what behaviors "might" occur in unobserved situations; (5) the immediate assessment of behavior during, or shortly following, observation; (6) the independent assessment of each specified behavior; (7) recognition and suppression by the observer of personal biases relative to individuals or behaviors; (8) care on the part of the observer to avoid such rating biases as the central tendency error, the leniency error, and so on; and (9) the replication of observations and assessments by independent, though similarly trained, observers.

It may be of interest to leave the research context for a moment to consider teaching observation and assessment in the "teacher evaluation" setting from the context of assessed teacher effectiveness in the schools. We shall try to summarize with a few statements relating the foregoing comments directly to teacher assessment as it may be engaged in for the specific purpose of judging teacher effectiveness. We refer to evaluation as it is employed in school practice for teacher selection, merit rating, promotion, and so on.

In order to assess teachers, and subsequently to evaluate them, we need to: (1) know what characteristics of teachers are amenable or available to observation; (2) decide upon a set of evaluative criteria—a set of values appropriate to the particular setting and often unique to the particular community or group (note that here we always are concerned with *value judgments*); (3) know something of the relationships of observable teacher characteristics to the criteria represented in the agreed-upon value system; (4) employ the most objective—reliable—available means of arriving at the assessments on which the evaluations will be based (usually involving the careful operational definition of the characteristics to be assessed plus training of the observers); and (5) employ advanced means of summarizing and synthesizing the assessments of teacher characteristics to generate over-all estimates that will be basic to the evaluation.

Returning now to some of the results of the teacher-observation phase of the teacher characteristics study, separate factor analyses of assessment for elementary and secondary school teachers led the staff to conclude that there were at least three major patterns of teacher classroom behavior that could be identified and therefore might be considered further. These were:

TCS pattern *X* warm, understanding, friendly versus aloof, egocentric, restricted teacher classroom behavior

TCS pattern *Y* responsible, businesslike, systematic versus evading, unplanned, slipshod teacher classroom behavior

TCS pattern *Z* stimulating, imaginative versus dull, routine teacher classroom behavior

These three behavior patterns were the major qualities of teacher behavior singled out for further attention. They were employed later in the study for making various comparisons of teacher groups. (Recall that these patterns were originally identified from the assessments of trained observers of teachers in their classrooms.) In another section, a self-report inventory—the teacher characteristics schedule—will be described. This instrument *also* provided estimates of TCS patterns X, Y, and Z (as well as estimates of seven other characteristics). However the X_{co}, Y_{co}, and Z_{co} estimates were derived, in that phase of the study, from those teachers' inventory responses correlating significantly with observer assessments X_o, Y_o, and Z_o.[1]

The Assessment of Additional Teacher Characteristics

The foregoing section has focused upon the observation and assessment of teacher classroom behavior. It has been noted that in addition to observation, both direct-inquiry and "correlates-type" data gathering instruments were constructed in the teacher characteristics study. In the first stage of development, these inventory-type instruments employed direct-inquiry items and were constructed by following the item analysis procedures and methods of validity determination familiar to psychometricians. These materials provided estimates of seven additional teacher characteristics:

TCS characteristic R favorable versus unfavorable opinions of pupils

TCS characteristic R_1 favorable versus unfavorable opinions of democratic classroom procedures

TCS characteristic Q favorable versus unfavorable opinions of administrative and other school personnel

TCS characteristic B learning-centered ("traditional" or "directive") versus child-centered ("permissive" or "indirective") educational viewpoints

---

[1] The subscript $_o$, as in X_o, is used to indicate data that were based upon observers' assessments; the subscript $_{co}$ is used to indicate "correlates" data yielded by inventory responses.

TCS characteristic *I* superior verbal understanding (comprehension) versus poor verbal understanding.

TCS characteristic *S* emotional stability (adjustment) versus instability.

TCS characteristic *V* validity of response versus invalidity of response.

A number of "opinionnaires" were developed in one set of studies. These related to teachers' attitudes toward groups of persons questioned in the school—the organization of teacher attitudes then was studied through factor analysis. Equipped with the results of the factor analyses, the study centered its attention chiefly on the attitudes of teachers (1) toward pupils, (2) toward "democratic" classroom activities, and (3) toward administrators, fellow teachers, and nonadministrative personnel (TCS characteristic R, TCS characteristic R_1, and TCS characteristic Q, respectively).

Teacher viewpoints with respect to curricular organization and scope, pupil participation and class planning, academic achievements standards, and so forth, also were investigated by means of direct inquiry and factor analysis of the intercorrelations among responses. This was done separately for elementary and secondary teachers. The factorial representation of educational viewpoints that emerged was not clear-cut; and there seemed to be justification for considering teachers' educational beliefs from the standpoint of a single continuum, rather than several factors. This variable has been oversimplified perhaps by designation simply as an "academic content-standards oriented versus flexible, permissive, pupil oriented" dimension (TCS characteristic B).

In order to obtain estimates of the verbal understanding of teachers, vocabulary and verbal analogy items were constructed, experimentally administered, and the responses analyzed. The procedure produced a small number of carefully selected items comprising a "verbal ability" scale (TCS characteristic I).

In a similar way, questionnaires were prepared (with such as, "Which of the following is more true of you? (a) I am more nervous than easygoing. (b) I am more calm than easily upset.")

These were analyzed to obtain items for use in estimating the emotional stability of teachers (TCS characteristic S).

A set of items intended to measure probable "validity of response" of teachers also was assembled. These were to be utilized in the detection of "tendency to make a good impression" when dealing with responses to direct questions.

The Teacher Characteristics Schedule

At a later stage of the teacher characteristics study efforts were directed at the derivation of correlates scoring keys that would permit estimation of the several characteristics from symptomatic responses to inventory items. This was done in order to provide more readily obtainable estimates of probable teacher classroom behaviors, and estimates of teacher attitudes, viewpoints, emotional stability, and verbal ability that were relatively less influenced by the tendency to give socially acceptable responses. Self-appraisal items were prepared, and (after a series of preliminary analyses) were included in successive forms—that is, sequential revisions, each subjected to rigorous response analysis against the criterion data—of a teacher characteristics schedule.

The teacher characteristics schedule was an omnibus self-report inventory based upon some twenty-five originally separate instruments. In its final forms, it consisted of 300 multiple-choice and check-list items relating to personal preferences, self-judgments, activities frequently engaged in, biographical data, and the like. (There was also an additional set of fifty items that provided control and group-identification information plus direct-inquiry estimates of verbal understanding, emotional stability, and validity of response.)

Three separate teacher characteristics schedule booklets were developed and employed—one for elementary teachers, one for English-social studies teachers, and one for mathematics-science teachers. Some items were common to the three booklets; others were unique to a particular booklet. Use of the teacher characteristics schedule made it possible to obtain estimates, from a teacher's responses, of each of the ten behaviors and characteristics, $X, Y, Z, R, R_1, Q, B, I, S$, and V.

Some Comparisons of Groups of Variously Classified Teachers in Respect to Estimated Teacher Behaviors and Characteristics

Comparisons of teachers in respect to their teacher characteristics schedule scores were replicated with two samples, one known as the basic analysis sample and the other as the survey sample. The basic analysis sample involved teachers, most of whom taught in relatively large cities, who were available for the systematic behavioral observations undertaken by the study. It was also used for inventorying in the development of correlates instruments. The survey sample, on the other hand, involved teachers from the United States as a whole, who had not been involved in the observation phase of the study and who had completed the teacher characteristics schedule upon the request of their respective principals (the sampling of principals having been made from national lists with systematic selection to provide probability samples of the available groups of schools).

It seemed reasonable to hypothesize that certain subpopulations might differ systematically from others in teacher characteristics. Therefore, comparisons were made not only within the group of "teachers in general" (the combined groups of elementary and secondary school teachers), but also within the elementary teacher group considered alone, the secondary teacher group alone, and within the major subsample of secondary teachers consisting of mathematics-science teachers and English-social studies teachers. A few of the general findings will be presented.

Teacher Characteristics in Relation to Early Experience

Certain characteristics of teachers may be traceable to behavior patterns that were expressed in related, but different, channels long before the individual entered teaching as a profession. For this reason, such childhood and adolescent activities as "playing school," "reading to children," "taking charge of a class for a teacher," and the like, were considered.

One finding of particular interest was developed from a comparison of the scores of those teachers who participated in *any* of the activities inventoried and those of teachers who participated in *none* of them. Significant t ratios (.05 level) were obtained between early participation in teaching-like activities and the following behaviors and characteristics: understanding, friendly teacher classroom behavior (pattern X_{co}); responsible, businesslike teacher classroom behavior (pattern Y_{co}); stimulating, imaginative teacher classroom behavior (pattern Z_{co}); favorable attitude toward pupils (characteristic R_{co}); favorable attitudes toward democratic classroom practices (characteristic R_{1co}); favorable attitudes toward administrative and other school personnel (characteristic Q_{co}); and permissive versus traditional educational viewpoints (characteristic B_{co}).[2] Teachers who said they had "read to children" and "taken class for teacher" generally tended to score higher than others. Participation in schoollike activities during childhood and adolescence thus may be of significance with regard to the present characteristics of teachers.

Teacher Characteristics in Relation to Age

There appeared to be little doubt about the existence of important differences between teachers in varying age groups with respect to a number of characteristics. Among sixty different F tests computed from the data for the teachers participating in the teacher-characteristics study, forty-five of the sets of differences between means were found to be *significant* at or beyond the .05 level. Generally speaking, scores of teachers fifty-five years and above showed this group to be at a disadvantage when compared with young teachers—except from the standpoint of pattern Y_{co} (systematic and businesslike classroom behavior) and characteristic B_{co} (learning-centered, traditional educational viewpoints). Since this was a cross-sectional rather than a longitudinal approach, it is questionable whether or not these age differences were dependent primarily on changes in the teacher as he or she grew older, or on college-encountered cultural influences—influences that are substantially different for the presently middle-aged teacher groups who entered their profession some thirty-five-forty years ago than for the young groups. On the other hand, it is also possible to hypothesize that the age differences can best be accounted for through an examination of whether teachers possessing certain characteristics remain in their profession while others not possessing those characteristics leave. In the absence of definitive answers it can only be recognized that age must be taken into account as a relevant independent variable whenever teacher characteristics are considered. In terms of personality, teacher variation with age interacts with the main effects sometimes described as contributors to teaching performance, and it may either make differences in such main effects appear important when they do not exist or obscure them when they are present.

---

[2] The subscript $_{co}$ used with a pattern or characteristic indicates that the data were obtained from "correlates" or inventory responses. The subscript $_{ci}$ indicates the data were yielded by criterion items, or direct-inquiry items, relating to characteristics I, S, or V. A subscript $_{o}$ used with a pattern indicates that data were obtained from observation of behaviors.

Comparison of the Teacher Characteristics of Men and Women

Men and women teachers in the elementary school appeared to differ with respect to only four of the personal-social characteristics studied. As a group, men scored significantly lower than women on characteristic Y_{co} (suggesting less responsible, systematic, and businesslike classroom behavior on the part of male elementary teachers). There is also the suggestion (t ratios significant at the .05 level in the basic analysis sample, and differences in the same direction, although not statistically significant, in the survey sample) that male *elementary* school teachers may be more favorable in attitude toward democratic classroom practices; more inclined toward permissive, child-centered educational viewpoints; and more emotionally stable than women teachers in the elementary school.

Differences between the sexes, often insignificant in the elementary school, were *fairly general* and pronounced among *secondary* school teachers. Women generally attained significantly higher scores than men on the scales measuring understanding and friendly classroom behavior (pattern X_{co}), responsible and businesslike classroom behavior (pattern Y_{co}), stimulating and imaginative classroom behavior (pattern Z_{co}), favorable attitudes toward pupils (characteristic R_{co}), favorable attitudes toward democratic classroom practices (characteristic R_{1co}), permissive educational viewpoints (characteristic B_{co}), and verbal understanding (characteristic I_{ci}). Men scored significantly higher with respect to emotional stability (characteristic S_{ci}) than did women teachers in the secondary school.

In breaking down the basic analysis secondary sample into subject-matter groups, differences between men and women English-social studies teachers were strikingly like those found in elementary teachers. Only two significant trends exist—for women to score higher than men in regard to responsible, systematic classroom behavior (pattern Y_{co}), and for men to score higher than women in regard to emotional adjustment (characteristic S_{co}). Among the basic analysis mathematics-science teachers, with a few exceptions, the differences fit the pattern found in the survey sample for secondary teachers in general.

Teacher Characteristics in Relation to Size of School

Considering elementary and secondary teachers combined, the means of teachers classified according to "size of school in which employed" differed significantly at the .05 level with regard to five of the characteristics studied: understanding, friendly classroom behavior (pattern X_{co}), stimulating, imaginative classroom behavior (pattern Z_{co}), favorable attitudes toward administrators and other school personnel (characteristic Q_{co}), verbal understanding (characteristic I_{ci}), and emotional stability (characteristic S_{ci}). The pattern is fairly clear. For these characteristics (as well as for Y_{co}, R_{co}, R_{1co}, and B_{co} where mean differences existed but did not attain statistical significance) teachers in large schools (seventeen to fifty or more teachers) scored higher than those from small schools (one-teacher schools, and three-to-five-teacher schools). The differences were most extreme with respect to verbal understanding (characteristic I_{co}).

Emotional Stability and Teacher Characteristics

Good mental health, or emotional maturity, generally is assumed to be a requisite for satisfactory teaching performance. Emotional instability and maladjustment are believed to be not only detrimental to the teacher's efficiency but also to be reflected in both the intellectual and emotional behavior of the teacher's pupils. Among elementary school teachers participating in the teacher characteristics study, emotional maturity (as estimated by the S_{co} and S_{ci} scores yielded by the teacher characteristics schedule) was correlated, moderately, with warm, understanding teacher behavior, with stimulating teacher behavior, with favorable attitudes toward pupils, with favorable attitudes toward administrators, and with measured verbal ability. Among secondary teachers, emotional maturity was positively (but not highly) correlated with favorable attitude toward administrators, favorable attitude toward pupils, and warm, understanding teacher behavior.

High mean emotional stability scores (S_{co} or S_{ci}) were attained by groups of teachers who, in making self-appraisals in other sections of the teacher characteristics schedule, (1) frequently named self-confidence and cheerfulness as dominant traits in themselves; (2) said they liked active contact with other people—as examples, would prefer supervising a summer playground to arranging books in a library or would prefer to direct games at a party rather than send the invitations, prepare refreshments; (3) reported their childhoods to have been extremely happy; (4) said they engaged in hobbies and handicrafts; (5) reported they seldom had been scolded or berated by parents when children; (6) expressed favorable attitudes toward other persons (as examples, confidence in intentions and efforts of their pupils' parents, belief that most people keep their emotions under good control); and (7) said they had engaged in occupations other than teaching for at least three years.

Lower emotional maturity scores were attained by teachers who (1) said their memories of childhood were unhappy; (2) seemed *not* to prefer active contact with others—for example, would rather send invitations for a party than direct games; (3) expressed distrust of intentions and reasonableness of parents and of their pupils; (4) were directive; and (5) were cautious in attitudes expressed toward their pupils.

"High" and "Low" Assessed Teachers Compared: An Approach to Teacher Evaluation

It will be of interest to review one of the study researches in which an effort was made to identify personal or social characteristics differentiating groups of teachers receiving uniformly high and uniformly low trained-observer assessments on the three major patterns of teacher classroom behavior (X, Y, and Z).

It may be emphasized that judgments of "good" and "poor" made about teachers are relative to the context in which the judgment is made—that is, the behavioral objectives of the teaching, the characteristics of the pupils taught, and the value system of the judge. This being the case, we should not be too optimistic about arriving at acceptable definitions for identification of groups of generally "superior" and "poor" teachers. However, one approach employed by the study provided data that may at least suggest clues regarding the characteristics of teachers many of us might be willing to call "good" and "poor."

An intriguing question arises regarding teachers who receive uniformly high, or uniformly low, assessments on *all three* teacher characteristics study patterns (X, Y, and Z) of teacher classroom behavior. In spite of the relativity of definitions, might a teacher whose composite assessment is (on *each* of three patterns, X_o, Y_o, and Z_o) one standard deviation or more above the mean—placing the teacher approximately in the top 16 percent of all teachers on *each* pattern—be considered a "superior" teacher? Might the teacher who is assessed one standard deviation or more below the mean on *each* of the three patterns be thought of as a "poor" teacher?

Regardless of whether or not uniformly highly and uniformly lowly assessed teachers may be considered "superior" and "inferior" (or "effective" and "ineffective") in a particular school system or community, such individuals can be thought of as prototypes insofar as the several kinds of teacher classroom behavior

Emotional Stability and Teacher Characteristics

Good mental health, or emotional maturity, generally is assumed to be a requisite for satisfactory teaching performance. Emotional instability and maladjustment are believed to be not only detrimental to the teacher's efficiency but also to be reflected in both the intellectual and emotional behavior of the teacher's pupils. Among elementary school teachers participating in the teacher characteristics study, emotional maturity (as estimated by the S_{co} and S_{ci} scores yielded by the teacher characteristics schedule) was correlated, moderately, with warm, understanding teacher behavior, with stimulating teacher behavior, with favorable attitudes toward pupils, with favorable attitudes toward administrators, and with measured verbal ability. Among secondary teachers, emotional maturity was positively (but not highly) correlated with favorable attitude toward administrators, favorable attitude toward pupils, and warm, understanding teacher behavior.

High mean emotional stability scores (S_{co} or S_{ci}) were attained by groups of teachers who, in making self-appraisals in other sections of the teacher characteristics schedule, (1) frequently named self-confidence and cheerfulness as dominant traits in themselves; (2) said they liked active contact with other people—as examples, would prefer supervising a summer playground to arranging books in a library or would prefer to direct games at a party rather than send the invitations, prepare refreshments; (3) reported their childhoods to have been extremely happy; (4) said they engaged in hobbies and handicrafts; (5) reported they seldom had been scolded or berated by parents when children; (6) expressed favorable attitudes toward other persons (as examples, confidence in intentions and efforts of their pupils' parents, belief that most people keep their emotions under good control); and (7) said they had engaged in occupations other than teaching for at least three years.

Lower emotional maturity scores were attained by teachers who (1) said their memories of childhood were unhappy; (2) seemed *not* to prefer active contact with others—for example, would rather send invitations for a party than direct games; (3) expressed distrust of intentions and reasonableness of parents and of their pupils; (4) were directive; and (5) were cautious in attitudes expressed toward their pupils.

"High" and "Low" Assessed Teachers Compared: An Approach to Teacher Evaluation

It will be of interest to review one of the study researches in which an effort was made to identify personal or social characteristics differentiating groups of teachers receiving uniformly high and uniformly low trained-observer assessments on the three major patterns of teacher classroom behavior (X, Y, and Z).

It may be emphasized that judgments of "good" and "poor" made about teachers are relative to the context in which the judgment is made—that is, the behavioral objectives of the teaching, the characteristics of the pupils taught, and the value system of the judge. This being the case, we should not be too optimistic about arriving at acceptable definitions for identification of groups of generally "superior" and "poor" teachers. However, one approach employed by the study provided data that may at least suggest clues regarding the characteristics of teachers many of us might be willing to call "good" and "poor."

An intriguing question arises regarding teachers who receive uniformly high, or uniformly low, assessments on *all three* teacher characteristics study patterns (X, Y, and Z) of teacher classroom behavior. In spite of the relativity of definitions, might a teacher whose composite assessment is (on *each* of three patterns, X_o, Y_o, and Z_o) one standard deviation or more above the mean—placing the teacher approximately in the top 16 percent of all teachers on *each* pattern—be considered a "superior" teacher? Might the teacher who is assessed one standard deviation or more below the mean on *each* of the three patterns be thought of as a "poor" teacher?

Regardless of whether or not uniformly highly and uniformly lowly assessed teachers may be considered "superior" and "inferior" (or "effective" and "ineffective") in a particular school system or community, such individuals can be thought of as prototypes insofar as the several kinds of teacher classroom behavior investigated by the teacher characteristics study are concerned. The description and analysis of the characteristics of such groups of teachers might be expected to provide potentially useful clues for the researcher, the teacher educator, and the administrator.

In the light of such thinking, the study staff undertook two investigations, one directed at the identification of characteristics differentiating between uniformly highly and uniformly lowly assessed teachers, and the other designed to assemble case data on a small, carefully selected group of elementary school teachers. Although independently conducted, the two studies complemented one another (at least as far as "generally highly assessed" elementary school teachers are concerned). In the second of the studies mentioned, *only* "highly assessed" elementary school teachers were considered; in the first, *both* "highly assessed" and "lowly assessed" and *both* elementary and secondary teachers were studied.

Reference will be made to the most comprehensive of these two researches; that is, where an attempt was made (1) to segregate criterion groups comprised of teachers receiving uniformly (a) high, (b) average, and (c) low observer assessments on *all* three major TCS patterns of teacher classroom behavior and then (2) to isolate characteristics that appeared to distinguish among the criterion groups.

The methodology involved will not be described in detail. Suffice it to say that with each teacher sample (elementary and secondary school teachers considered separately) the so-called "high" criterion group was selected so that it was comprised of teachers who received a composite observer assessment one standard deviation or more above the mean on *each* of the patterns X_o, Y_o, and Z_o. The "average" group for each sample consisted of teachers whose composite observer assessment was between 0.2 of a standard deviation below the mean and 0.2 of a standard deviation above the mean on *each* of the three patterns. The "low" group for each sample was made up of teachers whose assessments placed them one standard deviation or more below the mean on each of the classroom behavior patterns. Observer assessments represented the judgments of all observers viewing and assessing a particular teacher.

Among elementary teachers 4.5 percent of those observed met the criterion for the "high" group as against 2 percent for the "low" group. Similarly, with secondary teachers, 2.2 percent met the criterion for the "high" group, while 1.6 percent met the criterion for the "low" group.

It is appropriate to comment briefly on some of the differences between the uniformly "highly" assessed and the uniformly "lowly" assessed groups of teachers.

Scores on the teacher characteristics schedule scales that indicated favorable opinions of pupils, favorable opinions of "democratic" classroom procedures, favorable opinions of administrators and other school personnel, child-centered (permissive) educational viewpoints of philosophy, superior verbal understanding (comprehension), and emotional stability or adjustment showed the "highly assessed" group of teachers to be very clearly distinguishable from the "low" group. To a somewhat lesser extent, the "low" group was distinguishable from the "average" group and the "high" group from the "average" group in terms of the same characteristics. This was true generally for elementary and secondary school teachers alike.

An exception to the findings involved characteristic Q_{co} (attitude toward school personnel) where, while there was evidence of slightly more favorable attitudes toward administrators and school personnel on the part of the "high" group, differences were not large enough to have statistical significance. One other exception had to do with traditional-permissive educational viewpoints. Among secondary school teachers there was a distinct difference favoring the "high" rated group over the "low" group with respect to espousal of child-centered-permissive viewpoints. Among elementary school teachers, though the difference was in the same direction (favoring the "high" group), it did not attain statistical significance. The elementary teachers, who might be expected to be more permissive and child-oriented than secondary teachers, were somewhat more permissively inclined in the "high" group than in the "low" group, but the difference was not a large or a clear one.

With respect to the other broad areas of teacher personality, however, the highly and lowly assessed teachers, considered as groups, did stand apart. The so-called "high" criterion group, as compared with the "low" group, was (a) more favorable in its opinions of pupils, (b) more favorable in opinions expressed about democratic classroom procedures, (c) represented by a higher mean "verbal intelligence" score (this trend being more pronounced among elementary school teachers), and (d) represented by a mean inventory response suggesting superior emotional adjustment or stability.

These findings were supported by still another study in which a nation-wide sample of school principals was asked to nominate, anonymously, outstandingly "good" and "poor" teachers. The nominated teachers responded (again anonymously) to the teacher characteristics schedule. Separate analysis samples involving sixty elementary teachers, 138 mathematics-science teachers, and 151 English-social studies teachers showed numerous significant differences in inventory-based teacher characteristics scores. With regard to the elementary teacher sample, the outstandingly "superior" group of teachers attained significantly higher mean scores—compared with the notably "low" group—with respect to favorable opinions toward pupils, permissive educational viewpoints, verbal understanding, and emotional stability and also inventory estimates of friendly-understanding behavior, organized-businesslike behavior, and stimulating behavior. In regard to the mathematics-science sample, the "superior" teacher group was significantly superior to the "poor" group *only* with respect to verbal understanding and the inventory estimates of friendly-understanding, organized-businesslike, and stimulating behaviors. Among the English-social studies teachers, the outstandingly "superior" group rated significantly higher than the notably "poor" group with respect to favorable opinions of pupils, favorable opinions of democratic classroom behavior, favorable opinions of administrators and colleagues, verbal intelligence, and the inventory estimates of friendly-understanding and imaginative-stimulating behaviors.

These studies—different in orientation—have produced findings that are quite similar and in agreement.

It is interesting, also, to note some of the specifics (preferences, judgments, activities, and conditions) distinguishing this arbitrarily defined group of "good" teachers from teachers who were not so highly assessed. The "good" teachers almost uniformly expressed liking for personal contacts with other people—as indicated by such responses as: would prefer to sell stamps in a post office (rather than sort mail or collect mail), would prefer being a salesclerk (rather than handle complaints or work in a billing department), would prefer to test applicants for a position as typist (rather than compute typing test scores or maintain the typewriters used for tests).

Members of the "good" teacher group were very generous in their estimates of other people. They expressed belief that very few pupils are difficult behavior problems, that very few people are influenced in their opinions and attitudes toward others by feelings of jealousy, and that most teachers are willing to assume their full share of extra duties in the school.

The majority of teachers in the "high" group were between thirty-five and forty-nine years of age. They were married. They had made better than average grades while in college. They had belonged to high school and college social groups. They had received lessons in dancing, dramatics, art, or music outside of school as children.

These "highly assessed" teachers read a great deal. They had visited science museums and exhibitions during the current year. They frequently attended musical concerts and art lectures. Their activities suggested broad interest in cultural affairs.

On the other hand, characteristics that distinguished the "lowly assessed" teacher group suggested that the relatively "ineffective" teacher is self-centered, anxious, and restricted. This is indicated by responses such as: considers favorable prospect of professional advancement an important factor contributing to choice of teaching as a career; thinks a large proportion of teachers suffers from stomach trouble and ulcers brought on by unusual tensions related to their work; would train a pet dog as a watchdog (rather than teach it to respond to commands or to do tricks); feels a large proportion of people are inclined to worry more than they should; believes a substantial portion of parents' visits to school are made to criticize the teacher or the school; believes a fairly large portion of people (40 to 60 percent) are influenced in their opinions and attitudes toward others by feelings of jealousy; believes when teachers send students to the principal's office, it is more often the fault of the student than the teacher.

The majority of teachers in the "low" group were in the late middle and older age groups (over fifty years of age in elementary school teacher samples, over fifty-five years of age in the secondary school teacher samples). Among elementary school teachers, those in the "low" group were often unmarried. (This finding did not hold for secondary school teachers.) Teachers in the "low" group indicated preferences for activities that did not involve close contacts with people. With respect to certain kinds of experiences reported, the "low" group was almost the opposite of the "high" teacher group; that is, they read less, and manifested somewhat less interest in cultural affairs, science, art, and music.

These, then, were some of the findings relating to adequacy of teaching performance. Certainly the research has not settled the question, Who is the good teacher? However, there are some interesting suggestions here—some clues that may help to identify "good" and "poor" teachers if one is willing to accept the kind of definition employed in this research. Such a definition indicates that teachers are "good" if they rank very high among their colleagues with respect to such observable classroom behaviors as warmth and kindliness, systematic and businesslike manner, and stimulating and original teacher behavior.

Some Relationships between Teacher Characteristics and Pupil Behavior in the Classroom

Pupil behavior is the product toward which teacher behavior is directed. What were some of the relationships observed between the behavior of teachers and the behavior of pupils?

From the standpoint of the teacher characteristics study, pupil behavior was judged from observable acts of pupils in the classroom. Like the research on teacher classroom behavior, the approach was a direct one involving the observation and assessment of pupils in their regular classrooms and in the presence of their teachers.

It should be noted that this directly observed pupil behavior—behavior in process—may be quite different from the *products* resulting from pupil behavior. Some researchers indicate that it is more important and relevant to judge pupil behavior in terms of its products (for example, measurable skills, understandings, and attitudes acquired by pupils) than to assess samples of viewed pupil behavior. However, the method employed in the teacher characteristics study was a direct one and involved the assumption that pupils who carry on their classroom activities in certain ways will acquire intended learnings more successfully than pupils who behave or act in other ways.

The kinds of pupil behavior on which the study focused attention were:

"Apathetic-alert" pupil behavior
"Obstructive-responsible" pupil behavior
"Uncertain-confident" pupil behavior
"Dependent-initiating" pupil behavior

Assessments were made by trained observers in the same manner as were assessments of teacher classroom behavior.

No attempt will be made to go into detail regarding the various teacher-pupil behavior relationships that were studied. Three sets of investigations conducted by the teacher characteristics study will be summarized briefly: (1) one having to do with relationships between observed teacher behavior and observed pupil behavior, (2) another focusing on certain personal traits of teachers as they seemed to be related to pupil behavior, and (3) one relating to such conditions as grade level and subject matter taught by the teacher, sex, age, and marital status of the teacher in their effect upon pupil behavior.

Some conclusions may be drawn.

"Pupil behavior" appears to be rather *closely related* to "teacher behavior" in the *elementary* school. In the *secondary* school it seems almost *unrelated* to teacher behavior in the classroom.

In the elementary school, productive pupil behavior apparently is related to such teacher behaviors and characteristics as: understanding-friendly teacher behavior, systematic-businesslike teacher behavior, favorable teacher attitude toward pupils, favorable teacher attitude toward democratic classroom procedures, stimulating-imaginative teacher behavior, child-centered-permissive educational viewpoints, teacher emotional adjustment.

In the secondary school, productive pupil behavior seems relatively unrelated to all of the characteristics studied except "stimulating-imaginative teacher behavior."

Productive pupil behavior appears to be most evident in classes of elementary school teachers who are between thirty and fifty-four years of age, and in classes of secondary school teachers who are between the ages of forty and fifty-four. (Productive pupil behavior seems lowest in classes of elementary school teachers over fifty-five years of age, and classes of secondary teachers below thirty years of age.)

In the elementary school there seems to be little or no productive pupil behavior difference between classes led by men and those led by women teachers. In the secondary school, productive pupil behavior appears to be most evident when the teacher is a woman.

Productive pupil behavior in the elementary school appears to be about the same in classes of married and unmarried teachers (with a slight difference favoring married teachers). Similarly, among English or social studies teachers at the secondary school level, there appears to be no significant difference in pupil behavior when marital status of the teacher is considered. However, in mathematics and science classes in secondary schools there does seem to be a difference suggesting more productive pupil behavior in classes of single as compared with married teachers.

In the elementary school there appear to be no productive pupil behavior differences attributable to the grade level of the class (and the associated grade level taught by the teacher). In the secondary school, however, productive pupil behavior is high in social studies classes and low in English classes. It appears to be highest of all in mathematics classes taught by women teachers and lowest of all in English classes taught by men teachers.

Related Studies by Others

Although this chapter has dealt with the data yielded and the implications suggested by the teacher characteristics study, it is important for the reader to recognize other significant contributions being made concerning relationships between teacher and pupil behavior. It is not possible to summarize any large number of these studies.

We turn first to Gage's well-designed experimental study of pupil-observed change in teacher behavior. Comparison by Gage and his colleagues of experimental and control groups of sixth grade teachers indicated that the following occurred when teachers were provided with information as to how their pupils described their *actual* teacher (feedback) and their *ideal* teacher: (1) teachers' behaviors changed (as indicated by subsequent pupil description of their actual teacher) in the direction of the pupils' initial descriptions of their ideal teacher; and (2) teachers became more accurate in predicting their pupils' descriptions of "their teacher."

The work of Withall, Flanders, and others—following up and extending Anderson's exploratory studies, and the thinking growing out of Thelen's "Human Dynamics Laboratory"—have produced interesting findings on teacher-pupil interactions, particularly as pupil behavior appears to be influenced by "direct" (that is, "dominative" or "teacher-centered") and "indirect" (that is, "integrative" or "student-centered") teacher behavior. Flanders' research, for example, suggests that "indirect" teacher behavior does, indeed, foster increased pupil learning. However, those pupils who learned most and who scored highest on classroom attitude scales were those from classes exposed to *flexible* patterns of teacher influence (that is, predominantly direct teacher influence in certain situations—as when the student's perception of the goal was clear and acceptable—and predominantly indirect teacher influence in other situations).

The carefully planned and conducted studies of Mitzel and Medley have sought to detect relationships between identifiable behaviors of teachers in the classroom and accomplishments of the teachers' pupils—viewed from the standpoint of behavior change. Cook and his students from the University of Minnesota have been active in analyzing teacher attitudes to isolate pupil and teacher conditions with which such attitudes interact.

These are only a few of the investigations that, together with the teacher characteristics study, are yielding findings with some validity and are contributing to a better understanding of teacher behavior in relation to pupil behavior.

Implications of Teacher Behavior Research for Educational Practice

The reader may have found himself wondering about the importance of the research that has been conducted. He may have been speculating about the implications and applications of the research findings that are summarized. What are the restrictions, the limiting conditions, growing out of the methods through which the findings were arrived? Which must be kept in mind when proposals are made to apply such research to practical situations involving individual teachers and their education, recruitment and selection, assignment, supervision, or promotion?

Teacher Selection

Although selection, appropriately and realistically, is a continuing process extending over a number of courses, phases, experiences, and years, we shall deal with selection at two specific times —when individuals are selected for teacher education and when they are selected for employment. In both cases, of course, the success, or validity, of the selection process is judged by comparison of the selection data (records, interview data, test data, and so forth) with criterion data (data relevant to whatever definition and operational description of "expected teacher behavior" has been agreed upon). With this "ultimate" sort of criterion data in mind, and with attention directed to the prediction of such major patterns of teacher behavior as "understanding, kindly," "responsible, businesslike," and "stimulating, imaginative" behavior, what can be said that will help personnel officials responsible for teacher selection?

Our studies of long-time predictive or "presageful" validity have produced some low-order correlations between data relating to certain characteristics of teachers-to-be and criterion data relating to teacher classroom performance. The correlations are not high. While it appears that the predictors involved in the correlations *can* be useful when applied on an actuarial basis to large groups of teachers, their use can be expected to result in a substantial number of mistaken judgments (both false positives and false negatives) when applied to individual teachers.

Relatively low correlations in such prediction studies come as no surprise to the experienced researcher. The predictability of such a many-faceted criterion as that for complex behaviors is so limited by the problems of measuring both criterion and predictors (and also by practical conditions) that relationships representing common variance of perhaps one-fifth or one-fourth of the total variance probably approach the maximum expectancy.

With this expectation of relatively low predictor-criterion correlations in mind, let us consider what may be said about the usefulness of predictions made at teacher education and employment levels.

We can say, with a good deal of assurance, that predicting in-service teaching performance is possible to a greater degree when selecting teachers for in-service employment than it is when selecting for college (for teacher education). The longer the intervening time and the greater the intervening life experiences between the time of collecting predictor data and that of assembling criterion data, the less the probability of success in predicting criterion behavior. This is true for any kind of behavior predictions.

Regardless of the cautious and limiting statements made, there are many teacher characteristics that *do* merit attention as predictors of teacher behavior in a classroom. Some of these go rather far back into the life history of the individual. "Teacherlike experience" during childhood and adolescence, for example, represents one such predictor dimension. There are many others. We think we recognize what some of these predictors are now; we can identify still others that appear to operate but that still lack strong supporting evidence.

Teacher Education and Supervision

For practical purposes, "teacher education" and "teacher supervision" will be considered as two aspects of the same kind of activity—one preceding and the other following employment.

Here the same note of caution that was sounded in connection with teacher selection could be repeated. We have uncovered some clues in the form of correlates or symptoms of specified teacher performance. We are beginning to pull together factual information about teacher behavior that permits, for the first time, useful descriptions of the dimensions of teacher behavior and their interrelationships and interactions. This should be of significant usefulness to teacher education and to supervision in the schools—a taxonomy and analysis that make accurate description possible and that should lead to more definitive and operational methods for "acquiring teacher behavior" both in the teacher education program and on-the-job supervision.

There are, however, several important things to note. One is simply that the descriptions of teacher behavior obtained, regardless of how advanced the research is that aids in arriving at them, always to some extent reflect the guiding "philosophies" and the operating programs of teacher education institutions and in-service programs.

The descriptions resulting from teacher characteristics research are somewhat more refined and analytical than teacher education objectives, but basically they have been molded (sometimes almost accidently, it seems) by the philosophy of the education program to which the teacher was exposed—regardless of how vague, nonbehavioral, and nonoperational the statement of the program's objectives really may have been. We are saying, then, that the process involved in teacher behavior desciption and teacher education is a circular one, the descriptions of teacher behavior uncovered and the correlates of teacher behavior identified by research cannot tell what teachers ought to do or ought to be like. Research reveals only what teachers *do* do and what they are like in light of the education to which they have been exposed.

Also to be remembered here is that descriptions of teacher behavior begin with specific observed acts (phenotypical behavior) rather than the presumable underlying genotypical behavior. Specific acts or phenotypical behaviors are innumerable, varying with the situation in which the behavior has its setting. They probably are not highly transferable to other specific situations. Some intermediate "classes" of behaviors (for example, smiling by the teacher, verbal commendation, taking a pupil to lunch) probably are somewhat more transferable. However the inclusion in a teacher education curriculum or supervisory program of training directed at some of these fairly specific behaviors obviously would be ineffective in producing the more fundamental behaviors involved in teaching. It would be a great deal like coaching students for an achievement test—where learning is concentrated on specific items that really only sample the achievement content universe. Teacher education cannot be so pat. This limits the applicability to teacher education programs of certain findings of teacher characteristics research.

Finally, even if transferable teacher behavior is separated from specific teacher behaviors and attention is focused on the former, the question still remains of the degree to which the desirable behaviors are susceptible to training. Presently very little can be said about the extent to which certain personal and social behaviors may be altered. Obviously, teacher education should not neglect an effort to train for an important behavior simply because such an educational effort is difficult and time consuming; but there are practical limitations (of time and expense) that give pause when developing a teacher education curriculum.

Research of the teacher characteristics study type does suggest some specific correlates of teacher behavior, and also indicates how some of the specific behaviors tend to be intercorrelated to form aggregates or patterns of a more general sort. Teacher education may find use for some of the specific examples, but its proper interest will be in the transferability of behavior to various kinds of situations rather than in specific correlates as such.

Teacher Assignment

The actuarial nature of predictions in every area of human behavior must be kept in mind when teacher assignment is considered. When teachers are assigned to situations where they can be of greatest use, predictions are being made—predictions that sometimes are fairly accurate and sometimes miss completely. They cannot be right for each individual teacher, but for teachers as a group, more "hits" (successful assignments in the sense that the teachers' capabilities do fit the demands of the teaching situations) than "misses" can be anticipated. If not, the assignment system needs improvement.

Ideally, in assigning a teacher to a particular school or situation, an effort is made to match the abilities of a teacher to the situation in which teaching is to be done. In its grosser and very obvious application, this means assigning teachers educated for elementary school teaching to elementary school classes, secondary school English teachers to English classes, and so on. At a more refined level, successful assignment is judged in terms of the matching of teacher-behavior patterns, teacher viewpoints, and so on to the objectives of a particular school program and to the needs and requirements of the pupils to be taught.

In assigning teachers to school programs in such a way that teacher characteristics and school objectives are compatible, a great deal of success is possible once some of the important dimensions and patterns of teacher behavior have been identified. Some schools and school programs incorporate in their value systems strong preferences for "warm, understanding" teachers; others, while not disinterested in such "warmth" qualities, attach most weight to a "systematic, businesslike" approach in the classroom. Similarly some schools seek teachers whose attitudes and beliefs lean toward "permissiveness" and "child-centeredness"; others prefer teachers committed to "academic-centered" educational viewpoints.

Certainly, teachers as a group manifest large individual differences in any one of the variables just named. If teachers can be identified with such variables, the door to more successful assignment is opened. Teacher characteristics research definitely has suggested ways of identifying teachers with various patterns of behavior and teacher characteristics.

Matching teacher characteristics with pupil characteristics and needs complicates the problem (beyond the uncertainties involved in measuring teacher characteristics and pupil characteristics) by matters of administrative feasibility. It is one thing to identify the behavior patterns of teachers and to identify pupils who seem to have certain behavioral "needs"; it is another to match these successfully in the classroom. This demands not only knowledge of the essential relationships, but also attention to the important interacting problems of availability of teachers, administrative shuffling, scheduling, and so on. The administrative problem alone poses a formidable barrier to the assignment of teachers in the light of pupil needs. Theoretically, and viewed from the actuarial standpoint, it can be accomplished in the "ideal" school with fair success. (It should be added at this point that little or no research has actually tested hypotheses which predict successful teacher-pupil matching; the writer merely is stating his belief that such matching is possible once teacher characteristics are identified and estimated.)

Teacher Promotion and Merit Rating

Problems of teacher promotion and merit rating are too complex for adequate treatment here. However, (1) the process of teacher observation and teacher assessment can be refined to an extent that makes it practicable to attempt evaluation directed at teacher promotion and merit rating, but (2) such methods can be useful only after (a) the teachers and/or administration have carefully weighed and then stated explicitly the value system for judging teachers, and (b) the extent that reliable relationships between teacher characteristics, as predictors, and operationally defined values (criteria of teacher worth in the particular situation) have been studied empirically and determined pragmatically.

Limitations of the Findings

This chapter closes with a few summarizing statements about limitations and restrictions on applying the findings of teacher characteristics research.

1. The results that have been reported about the identifiable patterns of teacher behavior, the general and specific characteristics that seem to distinguish uniformly "highly" and "lowly" assessed teachers, relationships between teacher characteristics and pupil characteristics, and numerous other findings, all reflect merely what teachers are like here and now (circa 1950-1960, when the data were collected). The findings may be quite different at another time or in another cultural setting. In 1935 the picture

might have been a very different one. Five years, ten years, twenty-five years hence, teacher characteristics and their relationships may be changed significantly.

Generalizations in time (from one time to a future time) are possible only when the conditions are constant from the earlier time to the later. This is true in all science. In the physical sciences gross changes appear to come about more slowly. In the social or psychological areas, change is often relatively rapid. Therefore, descriptions and experimental findings applicable at one time cannot be assumed to be appropriate later unless additional study demonstrates substantial consistency over the temporal series—such consistency adding assurance, in the form of increased probability, to our inferences about future events.

2. The data obtained and findings reported concerning teacher characteristics do not indicate what "ought to be," nor do they imply that what has been found is "best" or "true" once and for all. Teachers who currently are rated "high" may manifest "permissive, child-centered" educational viewpoints (or some other pattern of teacher behavior) simply because their experiences have been oriented in this direction; that is, because they have been taught to adopt such attitudes or behaviors. In somewhat different words, research results of this kind vary, such variation depending to a considerable degree upon the cultural milieu providing the setting at the time and place of investigation. This is particularly true of those values and objectives prominent in the teacher training curriculum at the time the teachers under consideration were attending college.

3. Although this was mentioned only a few paragraphs earlier, the hazards involved in the application of most behavioral research findings to individual cases should be stressed again. For a group of 500 teachers, considered as a group and viewed actuarially, a principle of teacher behavior may be stated with rather high assurance; for the same behavior such a statement or prediction with regard to individual teacher A may be made with a probability only slightly better than chance and therefore with substantially reduced assurance.

4. The data assembled in the teacher characteristics study indicate that teacher behavior descriptions are not broadly appropriate to all teachers. Educational psychologists have learned to view the reliability and validity of tests as being relatively specific in nature (limited to the particular instrument used in a defined class for a defined purpose); we must maintain the same attitude toward the results of teacher characteristics research. Some of the findings do seem to be quite general and to permit state-

ments of principles that apply fairly broadly across different groupings of teachers. However, the manifestations of many teacher characteristics are specific to a particular teacher population. In line with this, it would be well to keep in mind that "effective teaching" probably differs from one grade level to another and from one subject matter to another.

5. The reported investigations of teacher behavior and teacher characteristics have been, to a large extent, of a taxonomic or descriptive sort and do not provide explanations of teacher behavior. This is true even for findings that pertain to relationships between various kinds of teacher conditions and teacher characteristics. For example, some of the results of the teacher characteristics study indicated significant differences in patterns of teacher behavior for teachers in different age groups. To describe this situation is one matter, but to try to "explain" it is another. Are elderly teachers different from younger ones, with respect to certain behavior patterns, because their behavior has changed as they grew older, or do the elderly teachers manifest different behavior patterns as a result of having been subjected to a different teacher education program than that of younger teachers, or, still again, do the differences result from "selective mortality"? These are questions that cannot be answered from cross-sectional, descriptive studies. We can say only that differences appear to exist. Of course, longitudinal studies could, and should, be conducted to clarify thoughts about such matters. However, lacking at present is the information on most teacher characteristics and their conditions necessary to permit the luxury of inferences about producer-product relationships.

Despite these limitations, evidence about teacher characteristics is accumulating—and will accumulate more rapidly in the future. Studies that are being conducted have great potential and will eventually be directly useful to teacher education and to practicing teachers and administrators.

4

R. L. TURNER

Teaching as Problem-Solving Behavior: A Strategy

Broadly considered, a research strategy is a three-stage affair aimed at solving a set of problems. In the first stage, the investigator develops a perspective on the problems he wishes to solve. This perspective includes those observations, assumptions, hunches, and problem definitions he expects to find useful in his movements toward his goal. In the second stage, the investigator adopts or devises research procedures and instruments consonant with his perspective. These two stages together comprise what may be called "strategy development." In the third stage of the strategy, the investigator finds out through empirical testing whether or not his strategy is useful in taking him to his goal and, equally important, whether or not it is useful in taking him beyond it toward new goals. This may be called the "utilization" stage.

This chapter was written not only to convey to the reader the results of using a particular strategy to gain better understanding of teacher behavior and proficiency, but also to suggest the process through which the strategy was developed.

Strategy Development

During the years immediately preceding 1957 (when work on the present strategy was begun), a great deal of research concerning teacher behavior and competence had been done but little systematic, useful knowledge had been acquired. In response to that situation, N. A. Fattu, in collaboration with the author, set about to develop a program of research into the cognitive aspects of teacher behavior and competence. Initially, the observation was made that experimental psychology—particularly the experimental study of learning—had been very productive during the same period the study of teacher behavior had been largely unproductive. This difference in accomplishment appeared to be due, in part, to the fact that experimentalists typically studied behavior in constructed situations where the stimuli were carefully controlled. Persons investigating teacher effectiveness or teacher behavior typically did not follow this procedure. The observation prompted the assumption that the study of teacher behavior would be more fruitful if situations with demonstrated relevance to teaching could be constructed.

In order to create such situations, an appropriate concept of teaching was needed. To this end, teachers were viewed as professionals specializing in the performance of a particular body of tasks or problems. These problems were thought of as involving the union of two broad variables: (1) instructional goals (behavioral expectancies) held desirable by some person or group, and (2) pupils at particular levels of development. This broad perspective of teaching implied that teachers engage in two central activities: (1) goal setting, and (2) instrumental responding to actualize these goals. The former was identified as problem-creating activity, the latter as problem-resolving activity.

After teaching was understood along the lines described above, it appeared to follow that the constructed situations might indeed be problem situations. These were thought of as being of two possible types. One was labeled "teaching tasks." From the time this type was first developed, "teaching tasks" has been understood to mean paper-and-pencil problems in which some aspect of a teaching situation is presented to the teacher. Typically, these tasks require the teacher to process considerable information about one or more pupils and to make decisions concerning goals or instrumental responses on the basis of this information. These tasks are short (usually taking ten to fifteen minutes to perform each), easily administered, and easily scored. They are useful primarily for studying problem-solving proficiency rather than the problem-solving processes of teachers. It was with these tasks predominantly in mind that this chapter was written.

A second type was thought of as a laboratory problem. Such a problem might involve many teaching activities of a simple sort or a single teaching task of a very complex sort. Typically, such problems require long performance times, are difficult to administer, and difficult to score. They are primarily useful in studying the problem-solving process of a limited number of teachers, as opposed to studying the problem-solving proficiency of a great many. These laboratory problems have been the concern of N. A. Fattu.

Once teaching was understood to be a problem-solving activity, it followed naturally that situations to which teachers were to respond should be problem situations or teaching tasks. In addition, there was possible an appreciation of the empirical relevance of constructed teaching tasks to teaching as well as a view of the connection between instrumental responding in teaching and instrumental learning as elaborated in academic psychology.

It should be possible to establish the relevance of constructed teaching tasks to teaching by observing how performance of these tasks varied among persons standing in different relationships to the actual practice of the professions. However, in order to specify with any clarity what these relationships were, it was necessary to subdivide the profession according to its specific practices. In making these subdivisions, the notion used was that teaching problems are involved with the intersection of particular behavioral expectancies and pupils at a particular level of development. Thus the profession was divided, as is characteristically done, by grade level taught and by instructional area; for example, teachers of intermediate-grade arithmetic and teachers of reading in the primary grades. A teacher may, of course, be simultaneously classified in several such subdivisions, as was the case with teachers in self-contained classrooms.

With the making of subdivisions within the profession, the relationships of various persons and groups to the profession became more specifiable. The group actually practicing within a particular subdivision became a criterion group. One such group was made up of teachers of intermediate-grade arithmetic. Other groups were then identified as peripheral; for instance, teachers of primary-grade arithmetic—who were seen to vary somewhat from the criterion group in terms of instructional goals and level of pupils taught. Secondary teachers of nonmathematical subjects were observed to vary still more with respect to goals and level of pupils. College-educated persons who did not teach at all were seen to vary widely from the criterion group, since the former was not formally concerned with any particular type of instructional goal or with pupils at a particular level. In general, it should be possible to demonstrate the relevance of a set of teaching tasks to teaching in a particular subdivision by showing that the performance level on these tasks varies with the "distance" between the formal concerns of the criterion group and those of the peripheral groups. The extent to which certain teaching tasks differentiate among these groups, of course, depends on their power to tap the specific concerns of the criterion groups. This power, in turn, depends upon how familiar the problem constructor is with the problems encountered in teaching in a particular subdivision and how much care is taken in constructing the problems. At a minimum, the problems should discriminate between teachers functioning within a given subdivision and comparably educated persons who do not teach in the subdivision or a related one or who are not trained to teach at all. If the teaching tasks did not make this distinction, there would be no means by which to show that they are problems particular to teaching, and hence, especially useful in talking about teachers and teaching.

The operational definition of "relevance" in the relationship between teaching tasks and teaching had definite implications. One was that not every problem a teacher might try to solve in the classroom could be viewed as a problem relevant to teaching—

relevant in the sense that teachers solve it better than any other group. Undoubtedly a substantial number of problems in the elementary classroom might be resolved equally well by teachers or comparably educated persons who do not teach. These problems may be viewed as making up the "common" or nonspecialized aspects of teaching, while those problems that teachers solve particularly well may be viewed as making up specialized or "uncommon" aspects of teaching.

The assumption that relevance was dependent upon the performance of teachers as opposed to the performance of other groups also had a second implication: teaching tasks could not be identified as "relevant" to teaching on a priori grounds. That a teaching task is relevant to teaching in a particular one of its professional subdivisions must always be shown empirically. By implication, it is likewise impossible to establish a priori the characteristics of the population of teaching tasks that compose teacher specialization. It follows that, at least at the outset, it is impossible to sample this population either on a random or a systematic basis. Thus, early in the process of establishing the relevance of teaching tasks to teaching, the researcher knows little about how these tasks represent the totality of teacher specialization even within a particular subdivision. Indeed, one of his problems, from a long range point of view, is to explore the parameters of specialization and to determine inductively special characteristics of the population of problems the performance of which will distinguish teachers from other persons. It is assumed that during this process the researcher might also be able to specify the characteristics of those problems composing the nonspecialized aspects of teaching.

As described above, the method of establishing the relevance of teaching tasks to teaching allows one to establish a minimal validity for these tasks as measures of specialized teacher behavior. It does not, however, explicitly set forth psychological grounds upon which to rest such relevance. That is to say, the method used actually does little toward identifying the underlying variables to which variation in teaching-task performance (within and between groups) may be attributed. These variables may, of course, be identified primarily by a "purely empirical" approach in which first one variable then another is related to problem performance. To a certain extent the author has used this procedure. However, the preferred strategy has been one consistent with the assumption that teaching is a kind of problem-solving activity. The essence of this approach is to infer, from the findings of academic psychology that bear on problem solving, the variables that should relate to the problem-solving performance of teachers.

At a very gross level, two such variables were identified. First was the general problem-solving ability of the respondent and his ability to benefit from learning experiences. This variable is related to general intelligence. General intelligence is not assumed to account for variation in task performance between teacher and nonteacher groups; it does account for part of the variation within these groups. If intelligence related to virtually all of the variation in task performance, it probably would be impossible to show that the tasks in question were relevant to teaching. There is no evidence to support a supposition that teachers are more intelligent than comparably educated nonteachers.

The second broad variable identified was the learning history of the performer. That is, the ability of a teacher to solve problems was hypothesized to depend less on such immediate matters as "insight," or the reorganization of the perceptual field, than to the extent to which the performer has had previous opportunities to acquire learning sets and specific responses relevant to the stimulus situations encountered in teaching. This hypothesis was based largely on the work done by Harlow (1949) and Osgood's *ad hoc* laws of transfer (1949).

Suppose, now, that the stimulus situations encountered are in the form of teaching tasks. At least three sources of variation in response may be identified as stemming from the learning history of respondents. First, the respondents may have had different degrees of opportunity to acquire learning sets and specific responses useful in dealing with teaching problems. Second, some respondents may have been improperly exposed to opportunities to learn to solve teaching problems so that they have acquired responses useful in avoiding such problems rather than in solving them. Third, some respondents may have acquired responses that interfere with their responses in problems of a certain type; that is, they mistake the problem stimuli for similar stimuli and give inappropriate responses (show negative transfer).

Among these sources of variation, the one dealing with different degrees of opportunity to acquire learning sets and specific responses has been most focused on the development of the present strategy. This source was chosen for attention primarily because it was thought that different degrees of opportunity to learn would account for a substantial part of the variation in performance.

With respect to amount of opportunity to learn to teach, it is apparent that the nonteacher, the person who is neither trained to teach nor teaches in a particular subdivision of professional practice, has had the least opportunity to acquire responses relevant to teacher specialization. Thus, with other variables controlled, the differences in teaching-task performance among teachers and nonteachers should be at a maximum (as was suggested earlier on other grounds). Correspondingly, as the number of opportunities to learn to teach increase, that is, as professional training and experience increases and other variables are controlled, differences in performance should become less. For instance, with other things equal, when performing tasks relevant to teaching intermediate-grade arithmetic, beginning education students with neither professional training nor teaching experience should perform least well, those with both courses in methods of teaching arithmetic and student teaching experience in the intermediate grades should do better, and finally, experienced teachers of intermediate-grade arithmetic should do best.

While particular learning treatments and experiences, such as those mentioned above, represent general opportunities for the teacher to learn more than the nonteacher about what is relevant to teaching, it is apparent that considerable variation may occur within a particular kind of learning treatment or experience. At the gross level, this variation in treatments and experiences may be identified with variation between preparatory programs in teacher preparatory institutions and variation between teaching situations in specific school systems.

Thus, variation in task performance attributable either to teacher preparatory programs or to teaching experience may be hypothesized to depend on *two* interdependent variables: (1) the specific instructional objectives stressed; and (2) the extent to which the teacher makes responses in situations involving these objectives. The relationships between these variables and their joint relationship to performance are unquestionably very complex. However, they may be hypothesized to be similar to the pattern outlined below.

Preparatory teachers enter a teacher preparatory program harboring certain instructional goals for pupils; that is, goals that have been reinforced by parents and previous teachers. Preparatory programs differ in the extent to which they modify these goals. Some programs modify them greatly; others, little or not at all. The degree of modification often depends on the extent to which individual instructors and associates elicit and reinforce *new* goals as well as the degree to which they alter goals already held by the student. Students from different preparatory programs, then, emerge from professional preparation equipped with varied instructional goals.

Going back to the notion expressed earlier that, in setting instructional goals for pupils, teachers create problems for themselves to solve, the hypothesis develops that when teachers from different preparatory programs first begin to teach they create somewhat different problems for themselves to solve because they hold different instructional goals. Hence, other things being equal, they will tend to acquire somewhat different repertoires of responses instrumental in solving these classroom problems.

However, other things are not in fact equal, particularly teaching situations. Within a teaching situation, a number of variables bearing on the instructional goals of teachers also operate: the expectations of principals, supervisors, and fellow teachers, each of whom can reinforce, not reinforce, or punish those responses that demonstrate the goals of the beginning teacher. Similarly the behavior of pupils may or may not reinforce the beginning teacher. This depends on the extent to which the teacher is capable of bringing about the behavior he expects from his pupils. The extent to which a teacher is successful in achieving his goals with pupils depends not only on what these goals are but also on whether he can perform the instrumental responses necessary to bring them about. Hence, his skill in problem solving itself affects the kind of goals chosen, and his problem-solving skills and his goals or behavioral expectancies become interwoven.

These variables operating within a teaching situation, in conjunction with those operating in the teacher preparatory program, may be hypothesized to have extensive bearing on the subsequent problem performance of teachers. First, if both the preparatory program and the initial teaching situation elicit and reinforce new instructional goals for the teacher, *and* the teacher has acquired a reasonably broad repertoire of responses prior to entering the classroom, he should be characterized by high problem-solving performance *and* the ability to enlarge the scope of problems he can solve. In other words, he will demonstrate professional specialization and moreover, will demonstrate a growing ability to deal with problems particular to the profession. On the other hand, if any of these conditions is missing, performance should be reduced. The extent to which it will be reduced must be viewed at this point as an empirical matter since we do not know exactly how these conditions interact.

There are, however, two possible cases or conditions of some interest. In the first case, the teacher finds himself punished in the initial teaching situation for using instructional goals heavily reinforced in the preparatory institution. This creates approach-avoidance conflicts. On the basis of Miller's conflict theory (1959), it may be anticipated that teachers experiencing this kind of situation will engage in vacillating behavior, probably leading to considerable classroom disorganization and some behavioral disturbance. The performance of such teachers on teaching tasks at the end of the first or second year of teaching would be expected to be poor. If overriding avoidance tendencies should be established, of course, withdrawal from teaching might also be anticipated.

The second case of interest occurs when the teacher does not expand his behavioral expectancies either as a student or as a teacher. This might result from poor motivation to teach or from inadequate opportunity to acquire new instructional goals. In any event, if such a teacher can respond instrumentally toward limited goals with pupils, hence gain reinforcement from pupils, his narrow range of goals should become very heavily reinforced. It follows that such a teacher will perform well only in a narrow range of problems, hence tend to perform poorly over-all when confronted with a wide range of teaching tasks. Moreover, it should be extremely difficult to change such a teacher since his dominant teaching habits, that is, his goals and instrumental responses, have tremendous strength.

In any case, the point of the matter is that the various interactions possible among the types of goals held by a teacher and those reinforced first by his college instructors and subsequently by his employers, associates, and pupils—these, together with his initial repertoire of instrumental responses (the repertoire available as teaching begins), create several different kinds of conditions (or cases). These conditions, in turn, orient the teacher toward his profession and ultimately determine his performance level. Thus, the years of teacher preparation, together with the first years of teaching experience, are a critical period in learning to teach.

Utilization of the Strategy

At this writing, four studies have been conducted relative to the strategy outlined above. Of these four, two have been conducted using teaching tasks in arithmetic and two using teaching tasks in reading. Of the two studies in teaching-task performance in arithmetic, the first was wholly exploratory, while the second was intended to verify and expand upon the results of the first study. The two studies in reading were largely exploratory—each utilizing different types of teaching tasks.

Teaching-Task Descriptions and Sampling Procedures

The initial investigation in arithmetic (Turner, 1960) utilized an instrument called "Form D, Arithmetic." It consisted of seven teaching tasks that appeared to hold promise (after informally trying out some twenty tasks in earlier forms—A, B, and C). These seven tasks were constructed to assess a teacher's skill in several kinds of performance: Two dealt with relating arithmetic materials to verbalized teaching objectives, two with diagnosing arithmetic difficulties among pupils, two with elaborating the meanings of the fundamental operations, and one with ordering arithmetic problems according to difficulty level for pupils.

These tasks were administered to 136 experienced teachers of arithmetic (grades three through six), to 185 undergraduates in elementary education, and to forty-one nonelementary school teachers. The group of 136 experienced teachers was drawn from two consolidated school systems, one having forty-five teachers in grades three through six, of whom forty-one were subjects, and the other having ninety-eight teachers in those grades, of whom ninety-five were subjects. The group of 185 undergraduates was composed of eighty students who had not yet taken a course in teaching arithmetic, a group of fifty-five students who were measured after completing the methods course and again after completing student teaching, and a group of fifty-two students who were measured after only student teaching. The group of nonelementary school teachers was composed of nurses, businessmen, housewives, and secondary school teachers. None had taught mathematics or arithmetic and none had taught in an elementary school. The nonelementary school teacher group had a slightly higher mean number of years of education than the elementary school teacher group.

Analysis of the data resulting from the use of Form D, Arithmetic indicated that several changes should be made in the seven teaching tasks used. Two of the problems were eliminated as unproductive; three of the remaining five were shortened and refined. These modifications resulted in a revised set of tasks called "Form E, Arithmetic." Because it is convenient to have two forms of a test, an alternate teaching task was constructed for each task in Form E. Five alternate tasks were combined to make up "Form F, Arithmetic."

Forms E and F were then used in a second investigation (Turner, 1961). The tasks were given to 101 Indiana teachers in grades three through six with one to ten years of experience and eighty-nine students in elementary education. The group of experienced teachers was drawn from five sources: thirty-two came from a township system in a city of about 500,000; twenty-one from a system in a city of about 18,000; nine from a system in a city of about 10,000; twenty-three were drawn from a summer school class at Indiana University; and sixteen were newly employed teachers from a system in a city of about 40,000. All teachers were assessed on Form E, Arithmetic.

Of the group of eighty-nine elementary education students, seventy-seven were assessed at the beginning of their course in arithmetic methods and again at the completion of student teaching, about eight months later. Half of this group received Form F for the initial test, half received Form E. All of these students, plus twelve who had taken arithmetic methods at other times, were assessed on Form E at the completion of student teaching.

The first investigation of teaching-task performance in reading (Wade, 1961) used nine teaching tasks developed by E. W. Wade (hereafter called "Reading Problems W"). Nine tasks were constructed to assess the skills of reading teachers in seven areas: (1) skill in selecting reading materials at the instructional level appropriate to individual children (two problems); (2) skill in grouping children on a criterion of oral reading accuracy (one problem); (3) skill in judging improvement in reading over a three-month period (one problem); (4) skill in recognizing the relative reading difficulty level of educational materials for elementary school children (one problem); (5) skill in diagnosing phonetic errors (one problem); (6) skill in diagnosing errors in structural analysis (two problems); and (7) skill in perceiving similarities among instructional reading exercises (one problem). Five of these problems were presented by tape recorder and four were printed problems.

These tasks were administered to 176 teachers of grades two to five, to ninety-seven sophomore students who had not yet taken a methods course in teaching language arts or any other methods course, and to eighty-nine students who had completed the bulk of professional undergraduate courses including language-arts methods and student teaching. Of the group of 176 teachers, 107 were drawn from a consolidated Indiana school system, and sixty-nine were nuns teaching in Chicago parochial schools. Of the 107 Indiana teachers, fifty-six had been assessed one year earlier on Form D, Arithmetic. Thus information was available on this group's performance in both arithmetic and reading. Of the eighty-nine students who had completed training in language arts and student teaching, eighty-eight were also assessed on Form E—thus the same information was available on this group.

The second investigation of teaching-task performance in reading (Burnett, 1961) utilized two complex problems calling for the diagnosis and correction of severe difficulties in reading. This device was constructed by R. W. Burnett. The two problems were administered to ninety-three teachers of reading, in grades one through six. Of these, forty-eight taught in or near a city of about 15,000 and forty-five taught in a metropolitan area of 500,000. They were also given to seventy-five undergraduates with no training in the teaching of reading and to nineteen Indiana reading specialists who met the International Reading Association's criteria for specialists.

Reliabilities

The split-half reliability of Form D was .84 among teachers and .87 among students. The median corrected coefficient of equivalence for the problems in Forms E and F was .88. The split-half reliability of the Reading Problems W was .85 for the teachers. The split-half reliability for the Diagnostic Reading Problems B was .33 for students, .76 for teachers, and .84 for reading specialists.

Empirical Relevance of Teaching Tasks to Teaching

In order to determine whether or not the seven teaching tasks in Form D, Arithmetic, were relevant to teaching arithmetic in grades three through six, the mean performance of the 136 teachers sampled was compared to that of the forty-one nonteachers with a *t* test. The differences were significant at the 0.001 level. As may be observed in Table 1, there was considerable overlap in the performance of these two groups: 10 percent of the nonteachers scored above the teacher mean. The tasks were accepted as relevant but in need of greater power. In order to increase their power, each teaching task in Form D was item-analyzed, and modifications were made. The modified tasks became Form E, for which an alternate form, F, was then constructed.

TABLE 1

Means, Variances, and Percentage Distributions of Teacher and Nonteacher Scores on Arithmetic-Teaching Tasks (Form D)

Group	*Mean*	*Variance*	*Percent in four score intervals*			
			4–11	*12–14*	*15–17*	*18–30*
Teachers	14.69	18.13	25	25	25	25
Nonteachers	10.88	11.60	54	36	8	2

In the studies following this initial one, the criterion for establishing the relevance of teaching tasks to teaching was modified to bring it in line with the later development of the strategy. That is, rather than selecting tasks that would simply distinguish teachers from nonteachers, the requirement was to distinguish among groups with widely different degrees of training and experience relevant to teaching in a particular subdivision of the profession.

Three groups were used to test the power of Reading Problems W to meet the above criterion. They were, (1) preparatory elementary school teachers with no relevant training and experience

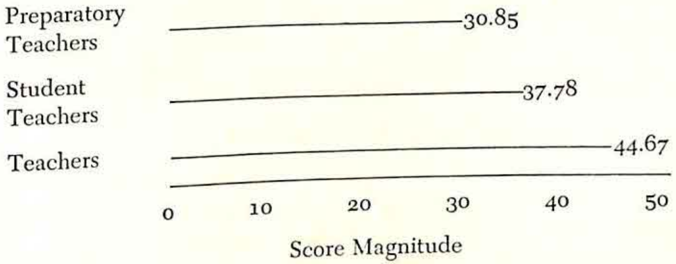


FIG. 1. Mean scores on Reading Problems W for preparatory teachers, student teachers, and experienced teachers. (Adapted from Wade, 1961.)

(who functioned in place of nonteachers), (2) graduating seniors in elementary education, and (3) experienced teachers of reading, grades two through five. Differences in the mean performances of these groups may be observed in Figure 1. Each mean differs from the other at the 0.01 level. These tasks were accepted as relevant to teaching reading.

Diagnostic Reading Problems B was designed to test more advanced skills in teaching reading than were tested by Reading Problems W. Thus, the relevant criterion groups were (1) preparatory elementary school teachers with no relevant training or experience (who functioned in place of nonteachers); (2) experienced teachers of reading, grades one through six; and (3) reading specialists. Differences in the mean performances of these groups may be observed in Figure 2. The difference between the means for preparatory teachers and experienced teachers was significant at the .05 level, the other mean differences at the .01 level. It is apparent that Diagnostic Reading Problems B is most relevant to the problems dealt with by the reading specialist and only of minor relevance to those dealt with by the average classroom teacher. To put the matter in a different way, classroom teachers show a comparatively low level of specialization in dealing with complex problems in teaching reading.

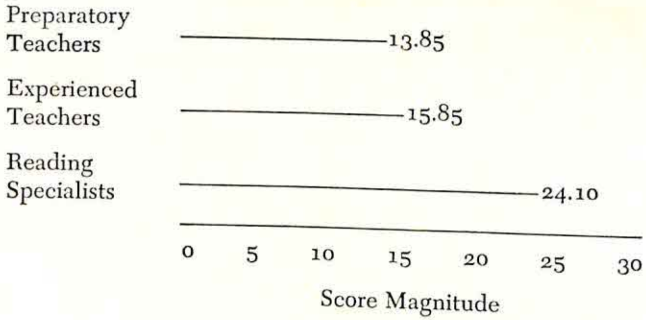


FIG. 2. Mean scores on Diagnostic Reading Problems B for preparatory teachers, for experienced elementary teachers, and reading specialists. (Adapted from Burnett, 1961.)

Relationship of Performance on Teaching Tasks to Measures of Other Behaviors

After establishing the basic relevance of teaching tasks in reading and arithmetic to their respective subdivisions of professional endeavor, the next concern was to determine the relationship between performance on these tasks and behavior on other types of measures. These measures were separated into two groups: those that measured either general intelligence or achievement and those that measured either attitudes or values.

For the general intelligence measure, ACE (Thurstone and Thurstone, 1952) raw scores were used. The correlation between ACE score and total score on each form, D, E, and F, for three independent groups of untrained, inexperienced preparatory teachers was .25, .32, and .40, respectively, and for three independent groups of graduating elementary-education students (one took Form D and two took Form E), .50, .29, and .45, respectively. On the basis of these correlations it appears that, on the average, intelligence accounts for about 15 percent of the score variance of teaching tasks in arithmetic. This relationship is considered to be in line with assumptions about the role of intelligence in task performance made earlier in strategy development.

The correlation between ACE raw scores and total scores on Reading Problems W among graduating elementary-education students was .56 ($p < .01$). The correlation between ACE scores and Diagnostic Reading Problems B among graduating elementary-education students was .02. The failure to obtain a significant r in the latter case was attributed to the extremely low reliability (.33) of Diagnostic Reading Problems B for this group (for whom the problems were too difficult).

The correlations between the coop C_2 test of reading comprehension (Davis and Davis, 1952) and teaching-task performance in reading and arithmetic closely approximated those between these tasks and ACE scores. This effect was attributed to the fact that ACE scores and coop C_2 scores correlated from .74 to .85 in the samples used. The correlations between arithmetic achievement, as measured by the STEP level 3, mathematics test (Educational Testing Service, 1957), and scores on Forms E and F for two independent groups of untrained, inexperienced preparatory students were .28 and .35 respectively. While the correlation between the STEP level 3 and ACE scores was not computed, it seems probable that a sizable proportion of the variance in arithmetic achievement scores is accounted for by intelligence.

In the main, the above results suggest that intelligence and achievement hold modest positive relationships to teaching-task performance, but that they by no means account for the total variation.

The relationships of teacher attitudes, as measured by the Minnesota Teacher Attitude Inventory (Cook *et. al.*, 1951), and of dominant interests in personality, as measured by the Allport-Vernon-Lindzey Study of Values (Allport *et. al.*, 1960) to teaching-task performance were investigated largely on an empirical basis. The results indicated that MTAI scores (indicating interests in interpersonal relations) were positively associated ($p < .05$) with performance scores on Form D Arithmetic among two independent groups of student teachers, but unrelated to performance scores on Form E among student teachers and among experienced teachers. MTAI scores also failed to correlate with performance scores in Reading Problems W. Correlation of each scale of the Study of Values (theoretical, economic, esthetic, social, political, religious) with performance on Forms D and E among independent groups of student teachers and among experienced teachers failed to demonstrate consistently significant relationships. There was no significant relationship between performance on Reading Problems W and any scale of the Study of Values. These results suggested that teaching-task performance in arithmetic and reading is largely independent of teacher attitudes or dominant interests.

Relationship of Specific Training and Experience Variables to Teaching-Task Performance

According to the theory developed, variation in task performance should be attributable not only to intelligence but also to specific learning treatments and experiences. Since the studies reported in this chapter were surveys of teaching-task performance, it was not possible to demonstrate that a change in performance resulted from a particular treatment or experience, but only that these two variables were associated in the histories of the individuals sampled.

Among elementary school teachers in preparation the treatments most relevant to teaching-task performance in reading and arithmetic were the respective methods courses and subsequent student teaching. In order to examine the effects of these treatments on performance using Form D, students who were entering a methods course in arithmetic were matched for sex and ACE score with students completing the course. No difference in performance between these groups was noted. The students completing the methods course were measured again after completing student teaching. There was no change in performance. These results were assumed to have occurred partly because Form D was a relatively insensitive measure, and partly because no design was used in which students served as their own controls throughout both methods and student teaching.

After the construction of forms E and F, the effects of an arithmetic methods course *and* student teaching, in combination, were examined again. Elementary-education students were measured at the beginning of the methods course and measured again at the completion of student teaching eight months later. For the purpose of statistical analysis these students were separated into two groups: those who did student teaching in a kindergarten, first-, or second-grade class, and those who taught a class in grades three through six range. As may be observed in Table 2, there was no significant difference between these groups on the pretest. Both made significant gains under the combined arithmetic-methods-student-teaching treatment. However, the group doing student teaching in grades three through six made a significantly greater gain than the group doing student teaching in the lower grades. These results suggest that arithmetic-methods-student-teaching treatment has a significant bearing on problem performance. They would also indicate that student teaching in grades three through six (to which the teaching tasks are conceptually most relevant) leads to significantly better performance than student teaching in kindergarten or grades one and two.

Two additional observations were made in the study of changes in task performance under the influence of the arithmetic-methods course and student teaching. First, gains in performance were not found to be correlated with intelligence. This suggests that, while intelligence is related to level of performance among student teachers, it does not account for changes in performance. Study was given to the possibility that the above result might have occurred because of low test ceiling. It developed that the top student teacher of the posttest was fifteen points below test ceiling; and the mean for student teachers in grades three through six on the posttest was five points below the mean for comparably trained, experienced teachers. This suggested that ceiling was not an important factor. In any event, failure to find a correlation between gains in performance and intelligence suggests that relationships between changes in task performance and variables unrelated to intelligence (such as motivational variables) might be profitably undertaken.

TABLE 2

Means and *t* Values of Pretest and Posttest Scores for Elementary-education Seniors Student Teaching in Either Grades K-2 or Grades 3–6 on Arithmetic-teaching Tasks (Forms E and F)

Group	*n*	*Pretest means*	*Posttest means*	*t*	*p*
K-2	38	20.92	24.50	3.340*	<.01
3-6	39	20.18	27.77	8.557*	<.001
t		0.045**	2.473**		
p		>.90	<.05		

* *t* test for correlated groups

** *t* test for independent groups

The second observation was that pretest task performance correlated only .41 ($p < .01$) with posttest performance—suggesting that ability to solve problems in teaching is not a stable trait during the last year of college. It seems an extremely plausible hypothesis that this trait remains unstable through the first years of teaching experience. If this hypothesis is valid, and there is some evidence that it is, the central problem in estimating future teacher performance is one of predicting the particular conditions or treatments under which teachers with particular personal traits will *change* in a given direction, rather than one of attempting to isolate traits upon which direct predictions to later teacher performance are made.

The relationships between teacher preparation and teaching experience among experienced teachers were anticipated to be extremely complex. In order to isolate at least the gross variables contributing to this complexity, the teachers responding to Form D were first divided into three categories according to the grade levels they had taught; grades three and four, grades five and six, and grades three to six. Statistical analysis indicated no significant differences in task performance by grade level. Next, these teachers were divided into groups employed to teach in rural, village, or city schools. Statistical analysis of the task performance of these three groups indicated no significant differences. Finally, the same teachers were divided into those who came from institutions of less than 1000 students and those from institutions of more than 1000. The difference between these groups was significant.

To examine the effects of years of experience on performance, the teachers responding to Form D were first separated according to size of graduating institution to control for this variable. They were then classified into three experience groups: one to three years, four to ten years, and eleven years and over. In order to have a group with no years of experience, graduating student teachers in grades three to six were utilized. Comparison of these groups from colleges with over 1000 students indicated that the significant rise in performance ($p < .01$) occurred between the "zero" year and the third year of experience and leveled off after the third year. This finding did not hold for teachers who graduated from colleges of less than 1000 students; this group performed no better than the student teacher group regardless of number of years of experience.

In follow up, by use of Form E, a number of changes were made. First, sampling was improved by obtaining experienced teachers from more diverse teaching situations. These teachers were classified into groups teaching in cities above and below 25,000 population. Second, approximately one third of the teachers sampled had never taught at a grade level above third. Virtually all teachers in the previous sample (using Form D) had taught a combination of grades (such as third and fourth), thus giving them experience in at least one intermediate grade. Third, the classification of preparatory institutions was changed from small and large to public and private. This change was made partly because all small schools in the samples were private and partly because school size is subject to increase from year to year.

Analysis of teacher performance according to these classifications indicated that third-grade teachers performed significantly less well than teachers of grades four to six ($p < .01$); that small-city teachers performed significantly less well that those from large cities ($p < .01$); and that teachers from private colleges performed significantly less well than those from public colleges ($p < .05$). In examination of the effects of years of experience on performance on Form E all third-grade teachers were omitted. Intermediate teachers were then separated according to type of preparatory institution, and the resulting groups divided according to years of teaching experience; one to two years, three to five years, and six to ten years. The numbers of teachers from large and small cities in these groups were proportional. For the "no experience" group, graduating student teachers who did their student teaching in grades four to six were used. Analysis of the performance of the teachers who attended public colleges indicated that the significant rise in performance occurred between the zero year and the end of the second year of experience ($p < .01$). Performance leveled off between the third and fifth year of experience. This finding did not hold for teachers who had graduated from private colleges; this latter group performed no better than did the student teacher group.

In exploring the training and teaching experience complex with Reading Problems W, it was not possible to determine whether or not the type of teaching situation was a determining factor since all experienced public school teachers were sampled from a single system. However, both types of preparatory institution and grade level were investigated. No differences were found among these teachers either by grade level or by college background. Teachers were then classified into four groups according to years of teaching experience: zero, one to three, four to ten, and eleven to twenty-five. Graduating student teachers were used as the zero-year group. Analysis of the differences indicated that a statistically significant rise in performance occurred between the zero year and the end of the third year of experience, and a statistically significant decline occurred between the tenth and twenty-fifth year. The latter result made it appear that teachers decline in ability to solve teaching problems toward the middle of their careers. However, analysis of the data obtained from the sample of nuns (who, typically, do not drop out of teaching) indicated that nuns with sixteen to thirty years experience significantly outperformed those with one to fifteen years experience ($p < .05$). There is, therefore, some evidence to suggest that the decline in performance among public school teachers as experience increases may be attributed to differential attrition from the profession rather than to other factors.

Exploration of the training and teaching experience complex with Diagnostic Reading Problems B indicated that grade level taught was not a significant variable. The difference between teachers from cities above and below 25,000 population was significant only at $p < .01$, while the difference between teachers trained at public and private colleges was significant at $p < .05$. It is also of considerable interest to note that reading specialists also differed in performance ($p < .01$) depending on whether or not they attended public or private institutions as undergraduates (with public college graduates performing higher). When the extent of additional training beyond the bachelor's degree for such specialists is considered, this finding is indeed startling. Inasmuch as no student teachers were measured, it was not possible to determine whether or not a rise in performance on Diagnostic Reading Problems B occurred with the first years of experience. Comparison of teachers with one to three and four or more years of experience indicated no significant differences in performance.

A question intrinsically related to the matter of teaching experience is, Does age, as opposed to teaching experience, influence performance? By holding experience constant at one to ten years and observing the performance of teachers in the three age groups of twenty to twenty-nine, thirty to thirty-nine, and forty to forty-nine years, it was possible to show on Forms D and E in arithmetic and Diagnostic Reading Problems B that, up to fifty years, age was not a significant factor in performance. Essentially what this finding means is that teachers who enter or return to teaching in their late thirties or early forties perform as well as those in their twenties. Whether or not teachers in their fifties and sixties perform less well *because* of their age remains, of course, an open question.

The over-all results of the investigation of training and experience variables suggest the following interpretations. First, there was considerable evidence that treatments such as methods courses and student teaching during undergraduate teacher preparation have a distinct bearing on teaching-task performance in arithmetic and in reading. Second, there was considerable evidence that variation in performance on teaching tasks in teaching arithmetic and on Diagnostic Reading Problems B in teaching reading is associated with variation in undergraduate preparation. Teachers from small and private colleges in the Midwest perform less well than teachers from large and public colleges in the Midwest. The fact that this result did not hold for Reading Problems W is significant because these teaching tasks were as closely associated with intelligence as were the teaching tasks for which the result did hold. This suggests that differences were due to variation in teacher preparation rather than to variation in student intelligence. Third, there is some evidence that variation in teaching-task performance is associated with variation in teaching situations. Whether or not this association is based more on selective employment practices than on differential learning treatments subsequent to employment could not be determined from the present studies. Fourth, there is considerable evidence that the very early years of teaching experience produce the greatest rise in teaching-task performance—as evidenced by differences in performance between fully prepared but inexperienced teachers and teachers with no more than three years of experience. There was little evidence to suggest that performance changed greatly, for the average teacher, after the third year of experience.

Relationship among Teaching Tasks

The correlation between Form D, Arithmetic, and Reading Problems W for experienced teachers (measured one year apart) was .50 ($p < .01$). The correlation between Form E and Reading Problems W was .28 ($p < .01$) for student teachers measured during the same sitting. When intelligence was not considered in the latter correlation, the resulting first order correlation was .04. The relationship between Reading Problems W and Diagnostic Reading Problems B was .14 (n.s.) for experienced teachers. On the basis of these correlations, it seems probable that there is considerable independence among the skills of elementary school teachers as measured by teaching tasks. What common factors there are appear to be based largely on general intelligence. There are no adequate grounds for the supposition that the typical teacher in the self-contained classroom will function on the same performance level in all the curriculum areas she teaches. It seems probable, in fact, that she may not perform evenly in a single curriculum area (such as reading) when diverse problems are encountered.

Current Research

Current research involving the use of teaching tasks in reading and arithmetic is focused on three questions. First, how is performance on teaching tasks related to other teacher characteristics, namely, those measured by Ryans' teacher characteristics schedule (Ryans, 1954)? Second, what kinds of changes occur in task performance during the first two years of teaching experience? Third, how are the changes in performance expected to occur during the first two years of teaching experience related to the conditions under which the beginning teacher gets her first experience, and how are they related to her characteristics as measured by the teacher characteristics schedule? If the author's current research yields coherent answers to these questions, it should be possible to suggest to school administrators some of the teacher characteristics and school conditions that will lead to better teacher performance at the end of the second year of experience. Further, it should be possible to indicate those characteristics and conditions that lead to static or, perhaps, poorer performance as experience increases.

Future Research

A very large number of studies is possible with the research strategy thus far developed. These studies fall largely into two groups. The teaching tasks presently in use represent only a very small part of the total number of problems with which elementary school teachers must deal. Moreover, there are, no doubt, several limits to the sensitivity of paper-and-pencil problems in assessing teacher problem-solving behavior. Thus, one group of future research studies should be concerned not only with a broader sampling of the population of problems encountered in teaching, but also with increasing the sensitivity of the measures employed (perhaps by the utilization of mechanical and electrical apparatus for presenting stimuli and recording responses). A second group of studies should be concerned with greater isolation of the sources of variation in performance of different classes of problems. Such research would be, in part, concerned with verifying the hypotheses already made in developing the strategy, refinement of those hypotheses already made, and development of additional ones. A series of studies is needed to lay the empirical basis for departing from strategy development by supplying quantities of systematically obtained data to serve as a basis for theory development. It may be observed in this connection that the data reported herein support some of the hypotheses made under the strategy —for example, that intelligence bears significantly on problem performance and that increasing opportunities to learn to teach in specific subdivisions of teaching practice are accompanied by increasingly higher performance on teaching tasks. However, these data serve only as indicators that the strategy is headed in a reasonably fruitful direction. They are at this point quite insufficient to support any but the most diffuse theoretical notions.

5

MILTON MEUX and B. OTHANEL SMITH

Logical Dimensions of Teaching Behavior

The study presented in this paper deals with the logical dimensions of teaching behavior and their effects on the response of pupils. The study grew out of an earlier investigation that attempted to determine whether or not instruction in the logic of high school subjects—history, geometry, English, and the sciences—would change the ability of students to think critically. In the course of the earlier projects, three questions arose that could not be dealt with adequately: (1) How can one describe the kind and amount of instruction given in the logic of the subjects? (2) What elements of logic are included in the instruction? and (3) What types of logical operations, if any, normally are performed by teachers and students in the teaching-learning process? The present research was designed in part to answer these questions.[1]

---

[1] This project is being supported by the U.S. Office of Education, Project No. 258 (7257), and is carried on under the direction of Professor B. Othanel Smith through the Bureau of Educational Research of the College of Education, University of Illinois.

Accordingly, the purposes of the present project are: (1) to devise a procedure for finding out whether or not there are logical dimensions of teaching and for describing such dimensions as may be found; (2) to determine whether or not teachers who are trained in the logical dimensions of instruction will exhibit teaching behavior that differs from teachers not so trained; and (3) to determine the effects of such different behavior as may be found upon the logical behavior of the students. In this paper are set forth the results of the first three years of the present investigation —the part that attempts to describe the logical dimensions of teaching behavior.

The Conceptual Framework

Teaching behavior may be studied at different levels, depending upon the purpose to be served by the inquiry. Perhaps the most elementary form of inquiry is called "natural history." At another level of investigation, an effort is made to discover the correlates of a given phenomenon. Inquiries of this kind are usually referred to as correlation studies. They are, typically, studies of static relations —studies requiring no deliberate variation of conditions. There have been many such studies of teaching behavior done mainly in order to find ways of identifying and predicting effective teaching. At a third level of analysis, the determinants of teaching behavior become the objects of inquiry. Rigorous studies at this level proceed by hypothetico-deductive methods—wherein definitions and postulates become the logical bases of the hypotheses to be tested experimentally.

The investigation of teaching behavior discussed in this paper is at the first, or natural history, level of inquiry. At this level, a phenomenon is observed, described, and classified. No effort is made to induce changes in the phenomenon, to study its underlying conditions, or to find its correlates. The modes of inquiry appropriate to this kind of research are those concerned with distinguishing among characteristics as well as those involved in the formulation of criteria defining the categories into which instances of the phenomenon are to be classified. One outcome of such a study is a classificatory system.

The conceptual framework used in a study of teaching behavior depends upon the level of analysis. If the aim of research is to determine the variables of which teaching behavior is a function, the conceptual framework will have reference to factors associated (1) with the genetic constitution of the teacher; (2) with his cognitive and emotional make-up; and (3) with the situation in which his teaching behavior occurs. On the other hand, when the aim of research is merely to describe teaching behavior, as in the present project, then the selection of concepts to be used depends upon the standpoint from which teaching behavior is to be observed and classified. Accordingly, the elements of our conceptual context are verbal behavior and the logical nature of behavior. Each of these will be discussed in turn.

Teaching as Verbal Behavior

The teacher performs a variety of acts in a range of contexts. He takes part in the activities of the school as a whole. He may become involved in school-community relations, and, of course, he is necessarily involved in carrying on the work of the classroom. Performance in all of these activities constitutes the generic category we call "*teacher* behavior." Although *teaching* behavior belongs to this general category, it should be distinguished from the more general concept. Teaching behavior consists of those acts that the teacher typically performs in the classroom in order to induce learning. Thus teaching behavior—as the term is used here —is to be distinguished from such classroom operations as giving examinations, grading papers, and keeping order. While these behaviors influence learning, they do not count as teaching behavior because they are not designed with the induction of learning as their explicit end.

Teaching behavior is primarily verbal. By verbal behavior is meant spoken and written discourse and symbolic expression (as in mathematical symbols). Of course, some teaching behavior is nonverbal. In the latter category are facial and bodily expressions associated with a teacher's momentary disposition—smiling, frowning, tone of voice, expression of eyes, and posture—and actions used to show how to do something, as when a teacher puts his fingers on a keyboard to show a student how to place his own fingers. However, observations indicate that teaching behavior consists chiefly of verbal behavior.

Skinner has treated verbal behavior at length (1957). Since his conception of verbal behavior is not the same as the one used here, it is important to distinguish his use of the term. For Skinner, verbal behavior is more comprehensive, including any movement capable of influencing another organism. Thus, speech is only one form of verbal behavior, although (as he points out) it is the commonest form. Skinner, however, also counts gestures—such as pointing and clapping of the hands and ceremonial acts—as verbal behavior, even though no vocal expression may be involved. To develop a general theory of communicative behavior, as Skinner undertakes to do, a comprehensive definition would seem to be appropriate. A problem in which a limited conception of verbal behavior is more appropriate here. Therefore, in this research the common-sense notion of verbal behavior is used, including only those forms of expression that use words or symbols. Gestures and other bodily expressions are grouped under the heading of nonverbal behavior.

Three types of verbal behavior are used in teaching. One type is characterized by the fact that the speaker intends it to have specific effects—it will instruct, elicit responses in some way, and be remembered. This type of discourse ordinarily involves certain intellectual operations, such as explaining and defining, and it sets forth propositions to be understood and restated on demand. The second type of verbal behavior is used in telling students how to perform some operation. The teacher may tell a student how to bend glass tubes, or how to improve his stroke in tennis, or how an assignment is to be done. This form of discourse achieves its purpose if the student is able to perform the actions in which he is instructed. The discourse itself is not meant to be remembered after the skill or operation is learned. The third type of verbal behavior consists of utterances in which the teacher praises and commends, disapproves and reprimands, advises, entreats, and enjoins. This kind of verbal behavior is commonly understood to have an emotional, rather than a cognitive, influence upon the student.

The Nature of Logical Behavior

Behavior, either verbal or nonverbal, can be said to be logical in two senses. First, behavior is logical if it can be formulated symbolically in logical form. For example, suppose an individual is given a number of slightly unequal rods to be arranged in order of their length. If he does so by looking for the shortest rod, then the shortest of those remaining, and so on—he constructs a series. His behavior—noting the rods and arranging them—may be described in terms of operations involving transitive, asymmetrical relations.

Further, behavior can be said to be logical if it can be modified progressively in response to logical self-monitoring—that is, self-analysis and correction of one's thought processes in accordance with rules of logic. This conception of logical behavior is not concerned with the structure of actual thought. Nor is it concerned with behavior in terms of logical forms. Rather, it is focused upon the governing of thought or of overt behavior. Both of the foregoing notions of logical behavior are used in this investigation.

In the following section are described two units of classroom discourse (the episode and the monologue), the criteria used to identify these units, and the reliability of these criteria. In subsequent sections are logical categories into which the episodes are classified, together with the reliability of the criteria for these categories. Finally, with regard to some of the more important categories, the logical structures within the episodes themselves are discussed.

The Unit of Classroom Discourse

It seemed clear that, for the purposes of this project, some kind of unit would have to be developed to serve as a means of breaking up and analyzing the mass and variety of verbal behavior contained in verbatim transcripts of classroom discourses obtained from tape recordings.[2] In this section is first a brief discussion of the requirements for a unit; then the two standard forms of the unit finally developed—the episode and the monologue.[3] The episode is a multispeaker unit; it focuses on the one or more exchanges that comprise a completed verbal transaction between two or more speakers. The monologue, a single-speaker unit, marks the "solo performance" of an individual addressing the group.

---

[2] The taping procedure involved using three microphones spaced evenly around the classroom, one of them close to the teacher. One project member operated the tape recorder and another served as an observer, recording im-

The Kind of Unit Needed[4]

For the purposes of the project, the unit of classroom discourse had to satisfy a number of requirements: a low inference level, logical analyses, neutrality with respect to subject-matter content, and a fair amount of reliability.

In order to ensure a low inference level, the unit had to be a simple behavioral one. That is, it had to consist of items of readily observable behavior, which could be described in simple, concrete terms; it had to focus clearly upon what teachers and students do and say in their sessions of instruction, recitation, and discussion.

These units of behavior had to be analyzable both in terms of the logical aspects of teaching operations and in terms of the thinking reflected in the classroom performances of the students.

By neutral unit was meant one that would take on the same general shape or "form" regardless of its "content." That is, a unit marked out in a geometry class transcript sould resemble, in basic form, a unit found in an English or a social studies transcript. This generality of form is one way in which a unit of classroom discourse can be neutral. It provides the common basis, continual

---

portant features of the context—such as text used, background material, and diagrams on the blackboard. The entire class period was taped for each class—instead of using a time-sampling method—because inadequate knowledge of context and of the population (set) of logical dimensions (their kinds, frequencies, and patterns) precluded developing a sound method of representative sampling.

[3] For a more extensive treatment of this aspect of the work, especially illustrations of episodes and monologues, see Aschner, Chapter III (1958).

[4] The remainder of this section is adapted from Chapters IV, V, and VI of Aschner (1958).

from class to class, subject field to subject field, upon which comparisons can be made.

A final requirement was that of reliability—a kind of unit that, ideally, would permit any analyst, trained to use the instrument, to identify the same units in a given set of data that any other analyst, equally trained and using the instrument, would find independently in the same set of data. How well the unit meets this requirement is reported below.

The Unit of Discourse

The two basic forms of the unit of discourse are the episode and the monologue.

EPISODIC DISCOURSE An episode is defined as the one or more exchanges that comprise a completed verbal transaction between two or more speakers. In its normal course, a discussion in progress exhibits a characteristic movement. We often hear this movement described in metaphor as "ebb and flow," "give and take," "talking back and forth," and so forth. Certain regular sequences of verbal interplay fall into endlessly repeated patterns of action and responding action.

The boundaries between episodes contained in the series of episodes constituting a discussion are traced, in part, by noting a number of conventionalized forms of speech and gesture that typify group verbal behavior. Certain standard forms of utterance are used, for example, to invite immediate reply; other forms are conventionally understood to forestall or prohibit immediate response. A direct question, addressed either to one person or to the group at large, conventionally demands some kind of responding action on the part of the individual or group addressed. A rhetorical question, on the other hand, is commonly understood to be uttered for its dramatic effect, as during a monologue. Although some do serve to trigger discussion, reply is neither expected nor invited. When a reply is made to a direct question, it is also conventional that the reply itself be acknowledged in some way—at least by word or gesture if not by further responding commentary or questioning. These and other standardized and conventional forms of verbal behavior fall together into definite patterns in the course of a discussion.

Two basic patterns of verbal exchange have been encountered in classroom discourse. The first basic pattern is called the "reciprocating pattern." Here is alternation between two speakers, with a regular back-and-forth movement aptly described as reciprocating action. The other basic pattern, called the "coordinate pattern," is one in which each successive speaker responds, more or less directly, to the entry rather than to the remarks of the immediately preceding speaker; hence each remark following the entry is coordinate with all the others.

Episodes are determined, not by shifts in speakers, but by shifts in speech forms, in what the speakers are talking about, or in how they deal with the point under discussion. For instance, one point of concern might be the identity of the discoverer of the Hawaiian Islands. A second episode might focus upon a different point—the date of the discovery. When this question is settled, the speakers might move on, in the third episode, to the point concerning the name first given the Hawaiian Islands. In each case, their dealing with the current point might go through a three-phase pattern: the point is raised in a question (opening phase); a reply is ventured and judged (sustaining phase); the exchanges are then either sustained further or dropped by a conventional affirmative (closing) remark. Episodes vary in many ways from this simple illustration. Nevertheless, each episode stands as a discrete unit, as a completed verbal transaction.

It should be noted, incidentally, that episodes do not always last through a final or closing phase. Quite regularly, one episode will be terminated without any explicit closing comment. Instead, the transaction is closed off by the advent of a new "entry": a remark advancing some new point that, in its own turn, touches off responsive talk. There is nothing out of the ordinary about a discussion that proceeds through such a series of "truncated" episodes.

When these patterns of exchange break down—fail to take coherent form on the verbal record—and when no single speaker is doing all the talking, it is likely that the discussion itself has become disorganized and confused. Other forms of scattered discussion occur in which everybody talks at once, and no one clearly seems to be speaking either to the point or to any other speaker.

MONOLOGUE DISCOURSE A monologue is defined as the "solo performance" by a speaker addressing a group. Monologue discourse does not reveal any great regularity of pattern or phaselike qualities as does episodic interplay. However, expository monologues in progress often exhibit a kind of paragraph-to-paragraph movement. A solo speaker seems, typically, to move from point to point in his exposition, raising one, elaborating at some length, and then proceeding to another.

Monologue discourse contrasts sharply with the back-and-forth interplay of discussion or conversation. The occasion and the conduct of monologue discourse are more strictly and more explicitly delineated by convention than they are in episodic discourse. It is understood that conversational exchange with the appointed speaker is neither expected nor invited. To address him, even after requesting and receiving permission to speak, normally is perceived to be an interruption, an intrusion. A monologue performance, like the episodic, is shaped by convention to stand as a discrete whole, a completed verbal transaction—in this case one undertaken between a speaker and a group. It is a transaction in the sense that speaker and audience have come together for a common purpose and that the group shall attend to what the speaker says.

Types of monologue activity characteristic of classroom discourse are didactic or expository discourse (the most common), announcements, and assignments by the teacher—presentation of assigned reports, recitations of memorized material, and so on, by the student.

Although there is variation among the classes studied, monologues occur much less frequently than do episodes. This is to be expected, since monologue performances are usually reserved for special purposes and special occasions in group discussions.

A Set of Criteria for Identifying Units of Classroom Discourse

The set of criteria used, on this project, for breaking classroom transcripts down into units of discourse is presented and discussed in Aschner (1958). It consists of a set of descriptive statements about verbal behaviors that can be observed during class sessions. The types of verbal action described are those that typically initiate, sustain, and close off the sequences regularly observable in the classroom. These units comprise the episodes and monologues described in the preceding section.

It is doubtful that the instrument in its present form would be useful to anyone lacking the conceptual and technical background that is presupposed on the part of the analyst. It was designed for use by a group of analysts who were already familiar with the conception of verbal behavior and experienced in the study of classroom transcripts.

Some features of the instrument, in its present form, should be noted. It begins with a brief list of definitions. Then, it presents four groups of criteria. Each group is numbered, and each criterion has its own number within a group. All criteria of the first group concern the identification of means of entry into episodes. Criteria in the second and third groups cover the kind of remarks found in the sustaining and terminating phases of episodes. The fourth group of criteria is used to identify monologues.

In marking off a transcript, the analyst tags each utterance with one or more numbers taken from the set of criteria. This number represents the identity, in his judgment, of the particular remark or remarks made. In the early applications of the instrument, the criterion numbers were always affixed to the utterances of any transcript. This practice not only facilitated testing—since disagreements in judgment were thus quickly located—but it also provided each analyst with some means of measuring and maintaining self-consistency in his own judgments.

It should be noted that the present form of the instrument is not considered to be a final one. Its formulation is still relatively crude at certain points and could be refined readily. It is likely also that additions and revisions may be made in the course of continued application of the criteria to transcribed discussions.

The Reliability of the Criteria

The type of reliability estimate used to assess the dependability of the criteria was a coefficient of interjudge agreement based on percentage of agreement between independent judgments. The experimental phase of the determination of reliability involved obtaining independent judgments of the utterances on the transcripts. The formula was based on the percentage of agreements of the total number of units marked.

Four project staff members were used to obtain the independent judgments required for the reliability test. These staff members had not participated in developing the criteria. However, it was very important that the judges be trained in order to acquaint them with the type of data being dealt with and, especially, to clarify the conceptual framework within which the judgments were to be made. The material used to train the judges consisted of actual transcripts of the class sessions. The training procedure for the judges involved studying the criteria, marking the training transcripts individually, and discussing difficulties with the staff members who had developed the criteria. The material used for the actual reliability estimates consisted of seven transcripts: three in eleventh-grade United States history and four in eleventh-grade English.

In the first step of final testing, each of the four judges marked off, independently, what he judged to be all of the units (both episodes and monologues) on each transcript. He numbered each verbal move on the transcript with the number of the criterion he thought appropriate. The judges were instructed to identify the greatest number of units that could result from the use of the criteria.

In the second step of final testing, the four judges worked in two teams. Within each team, the two judges considered together each verbal move in the light of the criteria and of their original assessments of each move. After this step, many disagreements were eliminated; others persisted.

Combined results from each team of judges were used in calculating coefficients of agreement.[5] The coefficients obtained are presented in Table 1.[6]

TABLE 1

Coefficient of Agreement for Identifying Episodes and Monologues Based on Selected Tapes from United States History and English Classes

Tape number and subject	*Coefficient*
1. U.S. History	.71
2. U.S. History	.73
3. U.S. History	.62
4. English	.71
5. English	.70
6. English	.64
7. English	.69

The coefficients range from .62 to .73—a fairly small range—with a median of .70.

The disagreements seem to stem mainly from the ambiguity of the verbal moves. Because of the many subtle shadings and nuances in behavior, utterances are not clearly specifiable as fitting one criterion or another. One consistently troublesome discrimination is that of applying criteria for distinguishing between a new aspect of the same topic and a clarification or amplification of some point in the topic. So far, it has not been possible to formulate a general and consistent rule to eliminate this difficulty. Another problem concerns distinguishing between long prefatory material and monologues.

---

[5] It was better to use the combined judgment of a pair of judges because of the complexity of both material in the transcriptions and criteria themselves. Both factors tend to produce a number of "oversights" that can be reduced by using pairs of judges to arrive at the final units.

[6] The formula used for the coefficient is: $R = \frac{A_{xy}}{Max(E_x, E_y)}$ where A_{xy} is the number of agreements between teams X and Y, E_x is the total number of episodes marked by team X, E_y is the total number of episodes marked by team Y, and $Max(E_x, E_y)$ specifies that the larger of the two numbers, E_x and E_y, be chosen.

Classification of Episodes

Episodes can be classified in a number of ways—by the nature of their content, by the number of verbal interactions they involve, by the psychological processes they entail, and so on. For purposes of this study episodes must be classified in terms of their logical features. In order to classify episodes in this way, two things must be done. First, categories must either be invented or those found in the domain of logic must be used. Second, criteria for deciding whether or not a given episode belongs to a particular category must be worked out. The development of categories is treated in this section.

Entries as the Base of Classification

Episodes are made up of three parts—an opening phase, a continuing phase, and a closing phase. As the task of working out a classificatory scheme for episodes began, a question arose concerning the part of an episode to be used—all or a particular segment. The opening phases could be classified and, thereby, the group episodes. By the same token, either the continuing phase or the closing phase could operate as the classificatory base. Or the episodes could be classified as such, without regard to their parts.

Certain considerations led to classifying episodes by their opening phases. The opening phase always contains a verbal move that evokes at least one, and more often a series, of related verbal exchanges. This verbal move is called an "entry." It is always a self-initiating move on the part of the person who makes it, and it is followed by responding remarks. The entry thus tends to shape the character of the episode. If the entry calls for the meaning of a word, the continuing phase is apt to consist of statements telling how the word in question is used. Likewise, if the entry calls for an explanation of something, the continuing phase will likely emphasize some form of argument or description. The words "apt" and "likely" in the preceding sentences were chosen deliberately. The continuing phase will not necessarily be consistent with the demands of the entry. The respondent (particularly a student) may misunderstand the entry, or he may not know how to respond and, thereby, fumble the verbal exchanges, or for some other reason fail to make an appropriate response. In such cases, there is not a close, logical correspondence between what the entry calls for and what the body of the episode contains. For this reason, the entry is a more dependable cue to the nexus of the episode than is the continuing phase of the unit.

Moreover, it will be of interest, later, to discover the extent to which discrepancies are allowed between what is demanded and what is supplied. Certain views of teaching emphasize permissiveness by teachers. This view assumes that a permissive atmosphere encourages student initiative, discussion, and creativity. At the same time, precision in the handling of ideas and symbolic operations is prized. It is socially—as well as pedagogically—important to discover the relationship between permissiveness and rigor in handling ideas. Were it to turn out that permissiveness and rigor of instruction are inversely related, readjustments in educational theory might well be demanded. The classification of episodes by entries would appear to facilitate the use of episodes in the study of teaching from this standpoint. One measure of permissiveness may be the extent to which the exchange of responses in the continuing phase is allowed to vary from the demands of the entries. Rigor of instruction, on the other hand, may be measured in part by the extent to which the handling of the body of the episode measures up to specified criteria of logic.

The Development of Categories

At the outset, it seemed possible to select ready-made categories from those found in general works on logic. If such selection were possible, the remaining task would have consisted of devising criteria and procedures for placing episodes in appropriate categories. However, it soon became clear that no such simple solution was possible.

The great variety and complexity of symbolic operations demanded by teachers made it clear that episodes could not be fitted neatly into ready-made classes. It was necessary to follow a more empirical procedure—to work out categories in terms of the nature of the entries themselves. Conventional categories of logic, such as definition, designation, and classification, have influenced the deliberations; and some of them appear in the list of categories. This is necessarily the case—for to ignore such categories would be to invent a new logic, and that task is not within our domain. However, the occurrence of these categories in the list resulted from studying the entries themselves rather than from a priori decisions.[7]

Beginning to examine a simple set of entries and to look at each entry as a whole brought an awareness that the logical character of an entry could be determined by reference to the sort of response it demands. This suggests that classification of an entry may be made by reference to the ideal response it demands.

The ideal response, of course, should not be confused with the actual response made by the student. It will be recalled that this latter was eliminated at the outset as a basis for classifying episodes because of discrepancies that sometimes occur between the demand of an entry and the character of the response. An ideal response, in contrast, is logically correct when the entry is not ambiguous.

The meaning of an ideal response may be further clarified by reference to examples of entries. The entry, "From which state did Mark Hanna come?" demands that a particular political unit be specified. It is not necessary to know the state from which he came in order to know that the entry requires, as an appropriate response, that a particular state be indicated. Nor is it necessary to know the time at which he came from that state. If an actual rather than an ideal response was requested, the ambiguity would need to be removed from the entry by a specification of time. Mark Hanna might have come from New York, or any one of a number of states, depending upon the occasion. In the sense of nativity, he could have come from only one state. Furthermore, an ideal response may be made in more than one way. In the present case, the particular state could be specified by naming it, by pointing to it on a map, or by sketching its shape, and so on. "Which line is the base of the triangle?" is an entry that likewise requires that a particular thing—a line—be specified. "What is the word

---

[7] Three additional approaches to developing categories based on entry verbs, entry nouns, and first words in the entry were studied and discarded before the approach now to be described was adopted.

(in the sentence) that is to be modified?" similarly requires that something—a word—be singled out. We can now generalize about what we have been doing and say that there is a set of entries that demands, as an ideal response type, that particular things be specified by naming or by pointing.

In contrast to the foregoing entries, are cases in which no particular thing is called for. In the entry, "What is some food material that the fish could use?" it is clear that food material is to be indicated. However, the phrase "some food material" is a variable. The entry, therefore, does not demand that a specific food item be indicated. Any of a number of foods could be named, each of which would satisfy equally the demands of the entry. In these cases, the ideal response type involves naming, or otherwise indicating, any one of a number of values of the variables satisfying the function of the entry.

Some entries demand more complex response types. "How did McKinley happen to be killed?" requires that a sequence of events leading up to and ending in McKinley's death be related. "How did they finally relieve this beleaguered garrison?" is an entry that also demands recounting a chain of events ending in the relief of the garrison. A similar response is to be made to the entry, "In the East, what had Cleveland done that made the capitalists unhappy?" Here, the response consists of narrating the acts of Cleveland that led to dissatisfaction among the financial leaders. In each case it is an ideal response type that consists of the narration of a sequence of events leading up to and culminating in some particular state of affairs. This state of affairs is said to be the result or outcome of the events, and the events are said to explain or to account for the outcome. Again, it is not necessary to know what the actual events were in order to tell that an entry demands the narration of events as a response.

Kinds of Entries

We will now list the categories into which the entries were grouped.

DEFINING Entries making up this group are concerned with how words or other symbols are used to refer to objects (abstract or concrete). These entries vary in form and content, but, in general, they ask implicitly or explicitly for the meaning of terms.

Defining may take the following forms: (1) A term is given and a definition or meaning of the term is to be supplied as a response to the entry; for example, "What does the word 'dorsal' mean?" (2) The entry contains neither the word "mean" nor the word "define." The entry asks instead what something *is*—for example, "What is a cablegram?" (3) The noun in the entry is a grammatically proper name. In these cases, the entry requires that the object designated by the proper name be described or otherwise indicated—for example, "Who was Paul Elmer More?" (4) Some entries ask for a term or expression that can be substituted for another term or expression—for example, "What is the symbol for gravity?"

DESCRIBING To describe is to represent something by words or drawing, to tell about something. Thus, the entries making up this category mention or suggest something and require that an account of this something be given. In the question, "What can you tell us about the gill rakers?" it is clear that we are asked to describe the gill rakers.

However, not all questions that mention or allude to something ask for a description. "What would be some examples of a sense organ?" is a question that names a class of things and asks that examples be cited. No description is requested.

In some cases, as in the example just given, it is easy to tell whether or not the entry requires a description or an identification; but in a large number of entries, this intent is obscure. "What did Cleveland find out?" is a question that might be answered by naming whatever it was that Cleveland uncovered, but the expectations would be more nearly satisfied were the question answered by a brief *account* of what he found out. On the other hand, "What is a common defect of this part (cerebellum) of the brain?" may be answered by naming the defect, although a description of the defect would not be inappropriate as an answer.

DESIGNATING To designate is to identify something by name—word or other symbol. The name designates the object (abstract or concrete) to which it refers. Thus, this group of entries is made up of items in which something is described or otherwise indicated, and the name used to refer to it or to identify it is requested. These entries vary widely in form and content. In general, they demand that objects (abstract or concrete) be designated by name or other symbol or simply by pointing. Consider the question, "What do you call a word used to modify a verb?" The question is answered by giving the name of the word—namely, "adverb." The question, "What reptile did he show in the film?" is answered in the same way—by giving a name—although the question does not explicitly ask what the reptile is called. Again, "What is the word (in a given sentence) that is to be modified?" is a question that can be answered by pointing to the particular word or by saying it.

Designating may take the following forms: (1) the entry demands that an example or instance, or a number of examples of a group of things be named; (2) the entry gives a set of things and requires that all members of the set be named; (3) the entry gives a particular class or group of things, or a particular object, and requires that it be specified by name or by pointing; (4) the entry describes or suggests something and asks explicitly for its name.

STATING Entries in this group do not ask for names, descriptions, and so forth, but for things to be stated. They may ask for statements of issues, steps in proofs, rules, obligations, theorems, conclusions, ideas, beliefs, promises, threats. For example, the question, "What is the conclusion?" asks for a statement of some sort. It seldom can be answered satisfactorily merely by naming or describing.

REPORTING The entries in this group ask for a report on what a book or document says, for information in the text, or for a summary or review, for example, "What happens in Chapter Ten?"

SUBSTITUTING The entries making up this category ask the student to perform a symbolic operation, usually of a mathematical nature, for example, "Substitute y for x in this equation."

EVALUATING To evaluate is to estimate the worth, dependability, or the like, of something. An entry of this type requires that some object, expression, event, action, or state of affairs be rated on its value, dependability, desirability, and the like. For example, the question, "Is he a good judge?" asks the student to rate a judge who acts in some particular manner.

OPINING To opine is to express beliefs, usually based on little or no evidence. Such beliefs are about what is possible, what might have been and is not, what might obtain in the future. "Do you think that historians will say that Wilson was right in proposing the League of Nations?" is an entry that asks for conjecture about how historians of the future will judge Woodrow Wilson with respect to a particular set of actions—those involved in proposing the League of Nations.

CLASSIFYING Each entry in this group makes explicit reference to an instance and/or class (group, set, kind) of things. The entry requires that a given instance be put in a larger class to which it belongs as a subclass. For example, "What special type of triangle did you find it to be?" is a question that makes reference, by the word "it," to a particular triangle. The student is expected to tell to what class of triangles this particular one belongs. As an illustration of questions that ask for a class to be placed in a larger class, consider the following: "What group of animals does the jellyfish belong to?" In this question, the term "jellyfish" does not refer to a particular jellyfish, but to a subclass. The student is required to name the larger class to which the group of animals called "jellyfish" belongs.

COMPARING AND CONTRASTING This type of entry requires comparison of two or more things (actions, factors, objects, processes, and so on). There are three kinds of entries in this category: (1) The entry merely asks for a comparison, the points of comparison not being explicitly indicated, for example, "What is the difference between probation and parole?" (2) The entry specifies two (or more) things and asks that either their similarities or differences be noted with respect to a particular characteristic, for example, "Is its (fish's) eye very large compared to the size of the grasshopper's?" (3) The entry names a thing and requires that another thing similar to it, or different from it, be indicated, for instance, "Which one (Canadian house) corresponds to the House of Commons?"

CONDITIONAL INFERRING This category consists of entries each containing an antecedent—that is, the conditional parts of a statement. In the sentence, "When it rains, the streets are wet," the phrase "when it rains" is the antecedent. The phrase "the streets are wet" is the consequent. Now, the entries that make up this category give an antecedent. Sometimes they give both an antecedent and a consequent. They never contain a consequent alone.

Here is an example of an entry containing an antecedent only: "How does that (undemocratic handling of colonies) affect the mother country?" The phrase "undemocratic handling of colonies" is the antecedent. It describes the condition of which the effect on the mother country is the consequent. The question asks the student to tell what the consequent is. Take another case: "If the diagonal (in a rhombus) is given as 12 and this angle is 60, what is the angle at *C* and at *A*?" In all cases where the antecedent alone is given, the entry requires that the consequent—effect, result, outcome, subsequent behavior—be supplied as the answer.

Consider an example of an entry containing both an antecedent and a consequent: "Did you ever get a headache from sleeping in a draft?" The phrase "sleeping in a draft" is the antecedent and "get a headache" is the consequent. In entries of this kind, the student is required to affirm the consequent, to deny it, or to say he does not know if he has ever suffered or enjoyed the consequent under the given condition.

Some of these entries ask for value judgments; some ask for statements of result or outcome; and others for descriptions of actions, decisions, and the like.

EXPLAINING There are several types of explanation entries, all of which have one thing in common: They give a particular consequent and they require that an antecedent be supplied. To explain is to set forth an antecedent condition of which the particular event to be explained is taken as the effect—or else, to give the rules, definitions, or facts used to justify decisions, judgments, actions, and so on. In the example, "Why did the light go out?" the consequent is "the light go out." The question asks the student to give a reason or reasons to account for the fact that the light is out. The reason(s) is (are) the antecedent(s).

There are six kinds of explanation entries, depending upon the kind of antecedent—mechanical, causal, sequent, procedural, teleological, and normative—used to account for the consequent. These are described as follows:

Mechanical-explaining entries give an event or action that must be accounted for through description of the way the parts of a structure fit or work together. For example, "How do fish make a sound?" asks for a description of the mechanism or structure that enables the fish to make vibrations.

Entries of the causal-explaining type give events, situations, or states to be accounted for and ask that a state of affairs be cited of which the given event (or situation or state) is taken to be the result. For instance, "What makes a person's muscles 'twitch'?" asks for a description of the condition of the nerves associated with twitching.

Sequent-explaining entries ask how something happened. They require that a sequence of events be cited of which the event to be accounted for is the end point. For example, as already cited, the question, "How did McKinley happen to be killed?" requires the recitation of events leading up to the assassination of President McKinley.

Procedural-explaining entries require description of the steps or operations by which a given result or end is attained. Here is a sample entry: "How did you get 72 (for an answer)?" The student is expected to tell the steps he took to obtain this answer.

The teleological-explaining type of entry contains descriptions of actions, decisions, states of affairs, or the worth of things, and requires that these be accounted for or justified by reference to purposes, functions, or goals. For instance, "Why are you doing those problems?" asks for a purpose—such as satisfying an assignment.

Entries of the normative-explaining type do either of two things. First, they may mention or assume a decision, judgment, or state of knowing and require that it be justified by citation of a definition or characteristic or both. For example, "Why do we call them (animals between vertebrates and invertebrates) the *chordata* animal group?" asks that the student give a definition of the *chordata phylum* and point out that the animals in question have the characteristics called for by the definition. Second, members of this group of entries cite actions, decisions, or choices (either made or to be made) and require that rules be given as reasons for the decisions, choices, etc. For example, "Why do we use 'shorter' (in comparing two pencils as to length)?" asks for a rule prescribing the use of "shorter" in such cases. Entries of this type usually call for grammatical or mathematical rules.

DIRECTING AND MANAGING CLASSROOM Many questions asked by teachers have little or no logical significance. They are not designed to evoke thought but to keep the classroom activities moving along.

Criteria for Classifying Entries

We turn now to the question of how entries to be classified are placed in the categories set forth in the preceding section. Two conditions must be satisfied in order to classify entries: first, criteria must be formulated by which to decide the category into which any entry is to be placed; and second, there must be tests of the degree of consistency with which different observers, using these criteria, put the same entries into the same categories.

FORMULATING THE CRITERIA The criteria were developed from a set of entries taken from a sample of transcripts covering all the subject-matter fields represented by the complete set of tapes. There were approximately 1400 entries in the sample set. Each entry was typed on a strip of paper for convenience in handling.

Two investigators, working separately, developed the criteria, largely through trial and error. One investigator worked through part of the sample set of entries, putting together entries that were similar to one another in some way. The gross experience of classifying entries in this way led to the formulation of a few rules. As other entries were classified, additional rules emerged. When this investigator had worked through the entire set of entries, the rules he had formulated were given to the second investigator who then attempted to classify the entries by the rules. The difficulties encountered by the second investigator were discussed, and the tentative criteria were modified to obviate the problems. By this exchange between the two workers, the criteria were put into semifinal form. The two investigators, working individually, now classified the entire sample of entries. They then compared their classifications, noted the agreements and disagreements, and made such changes as called for by their deliberations. The criteria thus revised were used as the working set.

RELIABILITY OF CRITERIA The type of reliability estimate used to determine the reliability of the criteria was a coefficient of interjudge agreement, based on percentage of agreement between independent judgments of a sample of entries. The formula for the coefficient was a simple one, involving only frequencies of agreements and disagreements.

EXPERIMENTAL PROCEDURE Four judges—graduate students in education from the University of Illinois—were used to obtain the independent judgments required for the reliability test. Two sample sets of entries were used for the purpose of training the judges. The first was an easy set containing about fifty-five entries. The other set was more difficult and contained about 120 entries.

Before classifying the first set of entries, the judges were given general instructions and a set of general procedures to be followed in classifying the entries. After classifying the first set of easy entries independently of one another, the judges met with the authors to check their agreement and to discuss difficulties they encountered. They then classified the second set of more difficult entries and afterward met with the authors again to check their agreement and to discuss difficulties they encountered.

The final set of about 300 entries was then classified by the judges. The entries in this final set were selected randomly from the entries judged to be in each category in the set of 1400 entries used in developing the criteria. In all but three of the categories—defining, explaining, and designating—one out of four entries was selected randomly. In each of the subcategories of defining and explaining, one out of three entries was selected randomly. In the subcategories of designating, one out of two entries was selected randomly. The one exception was substituting, which had so few entries that all were used in the final set.

Agreement on this final set of entries was determined through two pairs of judges. This procedure was used to minimize the number of sheer oversights likely to occur—rather than actual misjudgments and misapplications of the criteria. Once again, it was felt best that judges work in pairs rather than singly on this material, because its complexity tended to result in many oversights.

STATISTICAL INDEX AND RESULTS The formula used for the coefficient was based upon the number of agreements per category between the two pairs of judges. Thus each category has a separate coefficient.[8]

The coefficients obtained are presented in Table 2.

TABLE 2

Coefficients of Agreement for Classifying Entries According to the Logical Categories

Name of category	*Coefficient*	*Name of category*	*Coefficient*
Defining		Valuating	.60
(1)	.98	Classifying	.70
(2)	.88	Conditional Inferring	.67
(3)	.33	Explaining	
(4)	.00	(1)	.83
Total Defining	.84	(2)	.55
Describing	.67	(3)	.36
Designating		(4)	.67
(1)	.64	(5)	.67
(2)	.62	(6)	.76
(3)	.48	Total Explaining	.84
(4)	.90	Directing and Managing Classroom	.87
(5)	1.00	Opining	.73
Total Designating	.71	Comparing and Contrasting	.62
Stating	.63		
Reporting	.33		
Substituting	.88		

---

[8] The formula for the coefficient used is: $R_i = \frac{A_i}{A_i + D_{1i} + D_{2i}}$, where R_i is the coefficient of agreement for category i, A_i is the number of agreements in category i, D_{1i} is the number of entries placed in category i by the first pair of judges but not by the second pair, and D_{2i} is the number of entries placed in category i by the second pair of judges but not by the first.

As can be seen, the coefficients range from .00 to 1.00. The median is .67, and the middle 50 percent of the coefficients range from .62 to .84—a fairly high percentage of agreement for the present status of our categories.

Discussion and Interpretation

On the basis of the work we have done with the categories, it seems apparent that much of the unreliability is due to four distinct difficulties.

One is the overlapping between the categories due to the presence of conflicting cues within the entry; that is, one part of the entry indicates that it is to be placed within one category and another part of the entry indicates that it is to be placed in an entirely different category. Each of the two separate parts of the entry clearly conforms to different criteria. Thus, disagreement may result from a situation in which one judge follows one cue and another judge follows the other cue.

The second difficulty arises when neither the whole entry nor any significant part of the entry seems to satisfy the criteria of any category. That is to say, the entry does not seem to belong in any category. In this case, judges disagree over which of several categories might be most appropriate.

The third is somewhat similar to the first. It stems from what seems to be a gradual shading of one category into another, for instance, designating tends to shade into explanation, the entries seeming to fall along a continuum. Here, judges may differ as to what point on the continuum separates one category from the other.

The fourth difficulty arises from the fact that the criteria involve varying levels of inference—some extremely high in referential distance from the categories and others very low. The criteria for subcategories (1) and (2) in defining and (5) in designating require very little inference. These cases, as already noted, are among the categories high in reliability. It is interesting to note that both verbal and syntactical cues are strong and dependable in the criteria for these categories. The verbal cues "mean" and "define," as well as the syntactical cue "What is X?" go a long way toward indicating unambiguously the category to which the entries exhibiting them belong. On the other hand, some categories are marked by criteria requiring a great amount of inference. A case in point is subcategory (2) in explaining—causal explaining. The criteria in this case are not only vague, but also require the judge to decide if an entry asks for an explanation of an event or an explanation of an explanation of an event. Such an interpretation of a criterion involves a high order of reasoning, especially when neither verbal nor syntactical cues are dependable.

At present, we have been unable to develop completely independent categories or suitable criteria by which to eliminate such difficulties, and often it is not clear which need more improvement—the categories or the criteria.

Logical Analysis of Episodes

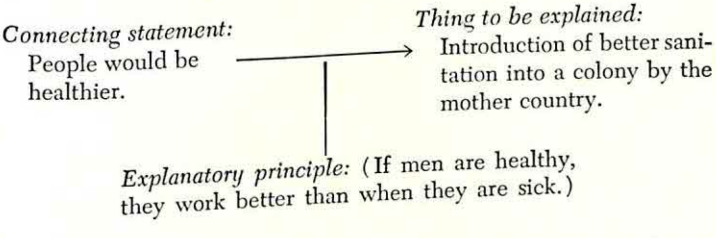
Verbal materials making up a curriculum are organized, in the course of teaching, according to the requirements of cognitive processes. They take on forms that can be identified and described. These forms give rise to the kinds of episodes identified by the criteria set forth in the preceding section. We shall now turn to a logical analysis of a sample of episodes selected to illustrate variations from an ideal logical pattern. Our observations indicate that teaching behavior seldom conforms to strict logical structures, such as those laid down in ideal forms of explanation, definition, and so forth. Of course, the question of whether or not teaching behavior should so conform is partly an empirical one, to be answered by reference to some criterion of effective teaching. Analysis of the logical structure of episodes, in terms of ideal forms, can indicate the kinds of variations in the logical dimensions of teaching and perhaps can suggest fruitful lines of research.

Explanatory Dimension

Some types of explanation—causal, normative, and teleological—have a common ideal pattern consisting of three elements: something that is explained, a principle of which the thing explained is an instance, and statements (usually factual) logically connecting the principle and the instance. Logically speaking, a complete explanation consists of giving the principle and the facts and showing that the thing to be explained is subsumed by the principle.

In some subjects, such a physics, it is possible to give a series of explanations in depth. For example, we can explain, in terms of Boyle's law, the increases of pressure in an automobile tire as it rolls down the highway. Boyle's law, in turn, itself can be explained by subsuming it under the molecular theory of matter. Such explanations in depth are seldom possible in subjects other than the basic sciences.

Consider, first, a complete explanation. A history class is discussing the effect on a country of holding colonies—and effects on the colonists themselves. The teacher asks, "Why would the mother country introduce better measures of sanitation?" A student says, "People would be healthier because of better sanitation." Then, the teacher adds the principle that healthy men work better than sick men. The pattern may be illustrated as follows:



Connecting statement: People would be healthier.

Thing to be explained: Introduction of better sanitation into a colony by the mother country.

Explanatory principle: (If men are healthy, they work better than when they are sick.)

Teaching behavior seldom exhibits complete logical explanations. Instead of complete explanations, we find episodes consisting of explanation sketches that either fail to give the connecting facts, or (as is more often the case) the explanatory principle. The following are typical cases of elliptical explanations in the classroom.

1. A passage from a chemistry text is read defining acids, bases, and salts. Examples of these are given, along with some of their commercial uses. It is pointed out that lead is used as a base in many paints and that lead is most often found, in nature, combined with sulphur as lead sulphate. After some further elabora-

tion the teacher asks, "Why does lead-base paint turn black?" A student replies that the lead combines (with sulphur). The pattern of explanation may be diagrammed in this way.

Connecting statement: It (lead) combines (with sulphur).[9] → *Thing to be explained:* Lead-base paint turns black.

Explanatory principle: (If lead in paint combines with sulphur in the air, it forms a black substance called lead sulphate.)

2. A class in history is discussing Cleveland's second administration. They are considering several unpopular things (at least, unpopular with certain groups of people) that Cleveland did during his second term in office. The teacher asks, "What did he (Cleveland) do that made the capitalists unhappy?" A student says that he repealed the tariffs. The pattern is as follows:

Connecting statement: He repealed the tariffs. → *Thing to be explained:* Capitalists are unhappy.

Explanatory principle: (If tariffs are repealed, capitalists are dissatisfied and unhappy.)

In each of these cases the explanatory principle is missing. Absence of an explanatory principle in an explanation sketch is not to be understood as indicating a poor explanation. There are, doubtless, many cases in which the principle is so clearly understood that to repeat it would be redundant. In still other cases, the principle is trivial and does not bear repeating. In some cases, however, the principle is at the heart of the explanatory process, and failure to make it clear is to give a faulty explanation. In dealing with the problem of finding the area of a rhombus where one angle is 60° and the shorter diagonal is twelve units long, the question arises as to how the class knows that an indicated angle is 120°. The student says that opposite angles in a rhombus are equal. If the principle were not given in this case, the explanation would be faulty because the correct principle is neither obvious nor understood from the context. Diagrammatically the pattern is as follows:

---

[9] Parentheses indicate what teacher and student assumed but did not say.

Connecting statement: (This figure is a rhombus.)

Thing to be explained: An indicated angle is 120°.

↓

Explanatory principle: Opposite angles in a rhombus are equal.

Valuative Dimension

In its ideal form, valuative dimension consists of four factors: (1) something—object, statement or expression, event, action, state of affairs—to be rated as good or bad, just or unjust, true or false, acceptable or unacceptable, and so on; (2) a rating; (3) a criterion or set of criteria by which the rating is made; and (4) facts about the thing to be rated that support the use of the criterion in making the rating. Suppose a class is discussing the question of whether or not certain statements that appear in so-called "true stories" are indeed true. To answer this question logically, the students must have a criterion for truth. In this case, the criterion they decide to use is that any statement conforming to observations is true. The set of statements appearing in the stories are to be rated as true or false. To apply the criterion to the statements is to get evidence that the statements are true—that is, that the statements correspond to observations. If there is such evidence, then the statements may be rated as true.

Valuation episodes, almost without exception, are elliptical. Either the criterion is not given, or the facts are missing. The following episode illustrates a valuative structure in which the criterion is not given. The class has been discussing the differential treatment accorded the Philippines, Cuba, and Puerto Rico after the Spanish-American War. The teacher asks, "Were Cubans really better prepared to take on the responsibilities of governing themselves than the other two groups?" A student answers that the Cubans were more prepared and adds that the Puerto Rican people did not have very much experience in self-government. In diagrammatic form, the structure of the episode looks like this:

Rating: Cubans were more (better) prepared. ← *Thing to be rated:* Preparation of Cubans for self-government in comparison to Philippinos and Puerto Ricans.

↑

Facts: The Puerto Rican people did not have much experience in self-government. (Supposedly the Cubans did, and nothing is said about the Philippinos.)

Criterion: Not given.

In the foregoing case, the criterion is entirely missing, and the facts on which to rest a value judgment are only partially cited. This example is typical of valuative episodes as they actually occur in teaching situations, although other elliptical forms often occur. In one of these, a rating is given without the justifying facts and criterion. In discussing a novel, the class is comparing one of the characters—Scobie—with Christ. The teacher suggests that Christ's weakness, from a human standpoint, was his goodness. Then she asks, "What is Scobie's main weakness?" A student answers, "It is sympathy (with mankind)." The teacher says, "All right." The logical structure of the episode is:

Rating: Sympathy (Scobie's main weakness). ← *Thing to be rated:* Scobie's characteristics as depicted in the story.

↑

Criterion: Not given.

Here, it is clear that neither fact nor criterion is given. There is no justification whatever of the rating—but only agreement between the teacher's judgment and the student's.

An infrequent incomplete form occurs when the evaluation criterion is given without any facts connecting the criterion with the rating. In discussing a novel, the teacher asks, "Is it fair for an author to use emotional appeal to promote an argument?" A student says that it is fair and then goes on to announce the criterion on which he wishes to rest his rating. The skeletonized form of the episode is as follows:

Rating: Yes. (It is fair.)

Thing to be rated: Use of emotional appeal by an author to promote an argument. Is it fair?

Facts: None given.

Criterion: Student A. If things appeal emotionally, then you get people interested.

Student B. Once you get people interested, then you can appeal to their reason.

Notes on Some Relations among Logic, Language, and Psychology

From the very beginning of the efforts to put classroom discourse into units and to classify them logically, a problem has stemmed from the fact that logical operations are expressed verbally in a number of ways. This is described below briefly. A second problem, faced from time to time, is the need for a broader conception of the psychology of both thinking and logic in the study of teaching behavior. In the long run, the former problem may turn out to be the more significant of the two, but it is nevertheless the one that the authors are less prepared to discuss at this time.

Logic-Language Relations

Language is used for many purposes. Some of these are clearly logical in character: to define a term, to carry on a course of reasoning, to prove the truth of propositions. Other purposes clearly are not logical in character: to give orders, to make promises, to express admiration. Still other purposes are neither clearly logical nor clearly without logic: to communicate information. Because language must be flexible and efficient enough to serve all these purposes, it does not achieve any one purpose as well as it might. Because of the multipurpose nature of language, some features of verbal behavior are puzzling when one is attempting to study logical operations expressed in verbal behavior. A number of distinct but related problems rise from such features of verbal behavior. They are evident when selecting and clarifying a unit of classroom discourse, constructing logical categories, and in making internal analysis of the logic of episodes. Presumably, they would have been easier to solve if the authors had had a complete catalogue of relations between logic and language in different contexts. Perhaps a complete catalogue of these relations cannot be developed. Nevertheless, it should prove interesting to note a few examples. Two have to do with logical operations and linguistic forms; and two, with the effects of context on the kind of logical operation called for by a sentence.

First, several logical operations can be called for by one kind of linguistic form. For example, "What is the horse?", "What is the reason?", and, "What is the method?" all have the form, "What is the X?" Yet, each calls for a different logical operation. "What is the horse?" calls for a definition or an identifying description of the animal. "What is the reason?" calls for an explanation or a justification of some act just done or described. "What is the method?" calls for an illustration or description of the method by which something is to be accomplished—making an object, conducting an experiment, and so on. Thus, it is clear in this linguistic form that a key word in a sentence—the word designated by X—determines the kind of logical operation called for by the linguistic form.

In other cases, one logical operation can be called for equally well by several linguistic forms. For example, "What does 'axon' mean?", "How is the term 'axon' defined?" and "The definition of 'axon' is what?" are all alike in one important respect. Each one calls for a definition of the term "axon." This is obvious, even though the terms and grammatical forms of the sentences are not alike.

In still other cases, the linguistic form clearly calls for a logical operation, with little influence of context. Examples of such sentences are questions like "What is the definition of imperialism?" and "Work is defined as the product of force and motion." Response to sentences of this type involve logical operations that conform closely to those taught in works on logic and philosophy of science. It is quite easy to identify the logical operations called for by such sentences.

A very different kind of example is to be found in sentences where the linguistic form does not clearly call for one logical operation, but may be interpreted as calling for any one of two or three logical operations. The kinds of logical operations required by these sentences usually cannot be decided uniquely without considering the context—which usually is sufficient to determine the kind of operation. An example is the question, "How would you define 'cultural lag'?" This can be taken either as asking for the definition of "cultural lag" or as asking for the method of defining to be used.

Logic-Psychology Relations

This study has been conducted as a natural history form of inquiry, involving observation, description, and classification of logical behavior in the classroom. The basic criterion of logical behavior used was *behavior governable by logical rules.* Thus, different kinds of logical rules formed the base for the different categories developed. This approach resulted in the following problems.

In developing the logical categories, the basis of each category was aimed toward providing the kind of rule a person would follow in forming an ideal response to the entry. Thus, in an ideal response to one kind of defining entry, the standard rules of definition would govern the response—rules such as making the definiens and definiendum the same scope, not repeating in the definiens what is in the definiendum (noncircularity), and not defining the term by what it does not denote (nonnegativity).

The responses to some entries, however, did not seem to be governable by rules. The opining and describing categories contain the best examples of this. For example, the response to an Opining entry such as, "What will the next generation say about President Eisenhower's administration?" does not seem to be governable by well-formulated rules. It seems to make more sense to speak of defining well than it does to speak of opining well. A similar point can be made about that part of the describing category concerned with describing or reporting one's feelings. What rules, for instance, does one follow in answering questions such as, "How do you feel about snakes?" and "Do you have headaches?"

The question then arises as to whether or not all our categories are logical ones—in the sense that the ideal response to an entry in the category can be governed by rules. Are some of the categories more psychological in character? How governable by rules does behavior have to be before it is considered logical?

The extent to which a student governs his response to an entry —by following a rule or a set of rules—depends on whether or not appropriate rules are formulated to answer the entry, on whether or not the student knows the rules, and on how well he can apply his knowledge of the rules in responding to the entry.

Even the available rules of definition—one of the best studied of logical operations—are not sufficient to govern completely the response to every situation calling for a definition. Since the main purpose of a definition is to make clear to the recipient of the definition what a term means, the teacher must judge the appropriateness of a definition he is about to give to the class. He must consider the age, knowledge, and other characteristics of the students before deciding which formulation of the definition is to be most effective for this particular class. There seem to be no well-formulated rules for every aspect of this procedure.

On the other hand, even in responding to opining entries, the student is guided by some rules (though they are quite general). For example, take the case of the opining entry cited above— "What will the next generation say about President Eisenhower's administration?" The rules to be followed in this case are those involved in examining what historians have said about presidents with beliefs and conduct similar to Eisenhower's, about presidents confronted with similar circumstances, and so on.

By the same token, responses to entries in the valuating category seem less governable by rules than defining entries and more governable by rules than responses to opining entries. A similar point can be made about the other categories.

It follows from the foregoing discussion that there is a rough continuum on which the categories can be arranged, a continuum defined by the extent to which an ideal response to an entry in the category is governable by rules. On this basis, it can be said that some of our categories are more logical than others. They are the categories containing entries whose responses are governable by rules to a greater extent than are responses to entries in other categories.

Finally, present narrow conceptions of the nature, scope, and method of both the fields of logic and psychology—and within psychology, especially the psychology of thinking—were hindrances throughout the investigation. The effect of these conceptions has been to preclude the availability of solutions to some of the most general and pervasive problems, solutions that might have been available under broader conceptions of these fields. Below are brief indications of how the present conceptions could be broadened, some examples of recent trends toward such broadened conceptions, and an idea of how such conceptions could have helped in this work.

Logic could be broadened in two ways—by investigating logical operations, such as defining, describing, and explaining, as they appear in ordinary language; and by investigating patterns of reasoning and argument used in various fields such as mathematics, physics, biology, and ethics. A trend toward such a broadened conception of logic is illustrated in the recent works of Toulmin (1958), Jensen (1957), and Hart and Honoré (1959).

Such a conception of logic would be helpful in identifying more accurately the kinds of logical operations that occur in the classroom. It would also help by providing a greater variety of logical forms—for example, kinds of argument used in everyday life—that could be used as standards to assess classroom discourse. In fact, Toulmin's work on the structure of practical arguments was quite helpful in the logical analysis of episodes, especially in the explaining and valuating categories.

The psychology of thinking could be broadened by (1) distinguishing and investigating different kinds of logical behavior versus behavior that is not logical; (2) determining the extent to which "rule following" is an accurate description of logical behavior, and the various kinds of rules that are followed; (3) determining the relative roles of rule following and imagination in various kinds of thinking, especially logical; and (4) identifying the various processes and abilities associated with performance and learning of various kinds of logical behavior (defining, describing, explaining, and so on).

Guilford's recent work (1959, 1960) illustrates a trend toward a broadened conception of the psychology of thinking, although it does not incorporate all the elements outlined above. For example, he has proposed psychological interpretations of logical behavior (such as deduction and induction). He has also included, as one of his operations, evaluation, which is similar in some ways to the authors' concept of logical behavior. And his products—especially classes and relations—are essentially descriptive concepts of logic. Unfortunately, however, Guilford's system could not be used because it does not include sufficient detail in the description and classification of the kinds of logical behavior found by the authors in the classroom.

A broadened conception of the psychology of thinking along the lines indicated—for example, the description of kinds of logical behavior in terms of kind of rules followed, if that should turn out to be appropriate—could have helped this work in three ways: (1) in developing logical categories for the episodes relating to the kinds of rules actually followed; (2) in analyzing ways in which the teacher monitors his own behavior and that of the students; and (3) in providing the basis for techniques of improving the teacher's monitoring of himself and of the students.

Logical Operations and Teacher Effectiveness

Teaching behavior is varied in style, tempo, strategies, and in the variety and order of logical operations involved in it. It leads, therefore, to a wide range of outcomes, both psychological and social. To attempt to determine the merit of any individual's teaching behavior is thus to deal with a many-sided problem, involving not only a large number of variables but also the question of what weight each variable is to be given in assessing the total behavior. This general problem cannot be dealt with in the present discussion. Certain considerations that should be taken into account as ways of determining teaching effectiveness are being worked out.

Perhaps the first task in the evaluation of teaching behavior is to determine response categories into which such behavior can be classified. A number of such categories, depending upon the standpoint from which teaching is viewed, can be constructed. For example, teaching behavior can be classified into such categories as information giving, approving and disapproving student behavior, making assignments, problem solving, and the like. Once categories of teaching behavior have been chosen, it is then possible to construct instruments by which to assess the behavior of the teacher in these categories and to measure the effects of such behavior upon students.

The analysis of teaching behavior provides a set of logical categories into which such behavior can be classified. While the question remains of what ends result from the behavior making up these categories, certain outcomes can be hypothesized. For one, it seems reasonable to suppose that, assuming other factors equal, students of teachers whose behavior measures high in logical operations would show higher scores on critical thinking tests than those with teachers having low ratings in logical operations. It would be expected that students of teachers who are superior in the handling of logical operations would rate high in the ability to identify mistakes in reasoning, in defining, in valuating, and in other logical processes. For another, improvement in the teacher's ability to handle logical operations would result in more student knowledge. The student would learn not only the usual facts taught in a subject, but, in addition, also learn the new relationships that proper performance of the logical operations brings out. The student would learn concepts, as he usually does, but he would also learn what a concept is and how to go about the task of clearing up vague and ambiguous uses of language when formulating and dealing with concepts. He would learn the laws of a discipline as he now does in physics. In addition, the student would also learn how laws are used in giving explanations of phenomena and how they are tied together in a system of definitions and theories. If increased knowledge and improved ability to think critically are acceptable outcomes of instruction and if improvement in the ability of teachers to handle logical operations results in such outcomes, then teaching effectiveness can be partly determined by reference to how the teacher handles such operations.

Teaching behavior can also be assessed by reference to internal criteria. Let us assume that clarity and preciseness of verbal expression are describable qualities of teacher performance. From a logical standpoint, the presence of these qualities is manifested by the absence of vagueness, ambiguity, and inconsistency. Now, the question of whether or not a teacher's behavior is clear and consistent can be answered in terms of how well he follows the rules governing performance of each logical operation. If a teacher handles all logical operations in accordance with the rules appropriate to them, it can be said that his teaching effectiveness is high. On the other hand, were the teacher's verbal behavior marked by vagueness and ambiguity and by inconsistencies in reasoning, it would seem reasonable to conclude that his effectiveness as a teacher is low.

Finally, the same behavior can be assessed in terms of other categories—for example, those developed by Flanders (1960) and by Ryans (1960). It might well turn out that teaching behavior judged to be highly effective in terms of the logical categories would be judged as ineffective in terms of other ways of classifying teaching behavior. A teacher might be effective in handling one logical operation and ineffective in handling another. His performance on a particular day might be judged as effective but, on another day, ineffective. Thus, the question of effectiveness is not easy to answer even after categories of teaching behavior have been set up.

6

PAUL V. GUMP

Environmental Guidance of the Classroom Behavioral System

For many years research on teacher effectiveness has focused upon the teacher's personal qualities in interaction with pupils. In this chapter is stressed a largely ignored alternative, the teacher's ability to develop an environment for effective learning.[1]

In order to communicate an understanding of this area, ideas and findings from three separate research endeavors are drawn together. The first research source is the work by Kounin and Gump investigating the "ripple effects" of teachers' disciplinary techniques; the second, the efforts of Gump and Redl to conceptualize and measure the influence of activity structures upon children's behavior; the third, the work by Barker and Wright in psychological ecology. The concept of the classroom as a manipulatable behavioral system might be furthered by use of yields from these three lines of research.

---

[1] This chapter was initiated at the Midwest Psychological Field Station, University of Kansas, and completed at the Education Research Project, Wayne State University.

The pupil in a classroom system is faced with a teacher, a group of age peers, an array of physical and temporal differentiations, and, finally, an activity pattern that weaves the elements of these classroom situations into a continual, unified environment. This is the context in which the pupil behaves and with which he must cope. The pupil does not only respond to discrete elements such as a teacher, or peers, or physical and temporal elements; his response is limited, supported, and coerced by a system that integrates these elements. His learning will be determined by the nature of the system and by his ability to utilize its potentials.

Research on the Ripple Effect

The interdependency of the elements in the classroom situation is illustrated by the research on ripple effect. This phenomenon occurs when an incident involving a teacher's control actions toward a deviant pupil has measurable effects on pupils who are not targets of, but audiences to, that teacher-deviant interaction. A review of ripple effect findings has been reported elsewhere (Kounin, Gump, and Ryans, 1961). Several studies that illustrate the interdependency of factors will be reported here.

The reality of the ripple effect was illustrated by a study of college students' reactions to punitive and supportive techniques directed by their instructor toward a classmate (accomplice) who arrived late to class. Students witnessing the punitive technique rated their instructor as less helpful, less likable, less fair, and less free from authoritarianism than did students viewing the supportive action. Experiments employing high school students and control efforts varying from punishing to ignoring yield similar results. Moreover, the students' rating of the seriousness of the deviant's behavior also was affected by the quality of the teacher's control technique—the more punitive the technique, the more serious the deviancy.

In the above researches, the single area of teacher's control behavior affected pupils' responses in several directions. The fact that other variables in control incidents affect individual pupils was illustrated by Gnagey (1960), who asked a "visiting teacher" to show films in natural classroom settings. Gnagey maintained a constant teacher-control technique but varied both the social power and the reaction of a "stooge" deviant. It was found that when the deviant child accepted the control technique in a submissive manner, as opposed to a defiant one, audience children rated the teacher as more fair, better able to handle children, and more expert on showing films. When the deviant was submissive, students also recalled more facts from the film. Effects were greater for boys than for girls and more pronounced if the deviant was of high social influence.

Interesting variations of the interdependency levels between a leader's control efforts and children who watched were discovered in a previously unreported inductive analysis of incidents of ripple effect in a boys' camp.[2] It has been suggested that every control technique to which an audience child responds adds something to that child's behavioral field. Audience campers used these additions in a number of different ways: they became more conforming, they gloated over the deviant's trouble, they drew conclusions about the counselor or the camp, and they amused themselves. Boys not only responded to the influence core of the deviancy control technique but to aspects extraneous to influence. For example, a counselor instructed a deviant to "stop wasting time and get cleaned up before the carnival." The cue "carnival" then stimulated the watching campers to ask questions about carnival tickets, activities, and food and drink.

Somewhat belatedly, it was recognized that control techniques can subtract elements from the behavioral field of the watching child. This subtraction involved a very different linkage between a control technique and an audience child. For example, a counselor (with no apparent intention of changing the audience's behavior) could remove a deviant, or an object, or an activity format that had been supporting the audience child's nondeviant behavior. With the support gone, the child had to cease or alter his behavior. Although this type of linkage is an obvious possibility if one happens to think of it, the author required the detailed data of specimen records before being stimulated in this direction. Before the careful study of these records, it had been assumed that *additions* to the behavioral field were the only source of the ripple between the control-technique incident and the behavior of the audience child.

---

[2] The study was conducted by Kounin and Gump (1958).

Those interested in further details on ripple effects should consult Kounin, Gump, and Ryans (1961). The point to be made here is that the interdependency among classroom factors is not a philosophical or theoretical assumption; it is an empirical fact with methodological implications. In the face of such demonstrated interdependency, the question arises: how can knowledge of impact of the total classroom system best be obtained?

Although it may be possible to perform tight experiments in which one holds all but one part of the system constant, and then vary that one part—this approach is difficult, time consuming, and artificial. Investigators of the ripple effect have just begun the task of exploring possible variations in factors influencing the ripple effect. It is apparent that not only are the specific techniques of the teacher important, but also the audience member's familiarity with the situation, his liking for the teacher, his motivation to learn, his sociometric linkage to the deviant, his behavioral linkage to the deviancy event, and so forth. In order to achieve a reasonable grasp of the complex classroom situation, one requires concepts and methods that either compare entire systems or changes in one large part of the system to changes in other parts. The isolation, manipulation, and measurement of very specific variables, as in traditional experimental psychology, may be neither feasible nor fruitful.

Research in Children's Activities

Investigation of large system variables was undertaken by Redl and Gump in a series of studies concerning the influences of recreational activities on children's behavior. Their approach was ecological rather than clinical. The usual procedure in clinical work with children is to view activity settings as arenas in which children display their particular personality patterns. A generally passive role in determining behavior has been assigned to the activity patterns and props of such recreational devices as painting, games, crafts, outings, and swimming. This investigation took seriously the possibility that such settings significantly influence the perceptions and behavior of participants—sometimes in fashions unique to the individual, at other times in ways applicable to the total group. The concept of *activity setting* and the allied concepts of *pattern of constituent performances* and *respondent behavior* were proposed. The referents for these concepts can be illustrated by a consideration of children's games.

A child's game can be labeled an activity setting. The word "setting" refers to the physical and temporal supports and limits involved in the play. A game covers certain time spans and employs a site with appropriate differentiations, enclosures, and props. The word "activity" refers to the game-given patterns of action. In general, an activity is supported by the setting.

In real game play, an observer can see two distinct aspects of behavior. He sees, on the one hand, a number of game-required behaviors: these are patterns of activity that come from the game itself. At certain points in a baseball game, a pitcher will throw a baseball toward a catcher and near a batter. He does this—not because of any unique need, personality trait, or mood—but because the game requires it. He is following a pattern of behavior accepted when he entered the game. Games lay down very specific and integrated action formats for their players. Rules of the game specify how positions are to be played, how props are to be used, how—in an action sense—opponents and allies are to relate to one another, what physical activities are legitimate, what areas can be entered, and so forth. When children play out these "requirednesses" of the game, they are engaging in the game's *constituent performances.*

Looking at the game play further, the observer can see a second aspect of behavior that is not game-required. There are the phenomena of pleasure and displeasure, of triumph and discouragement, of pride and shame. Arguments develop if the game makes possible ambiguous decision points: was the ball over the plate or a few inches wide? The game generates temporary attitudes of affiliation and contention. Many varieties of behavior beyond the constituent performances are observed in games. These behaviors are the *respondent behaviors.*

Some experimental evidence and a good deal of practitioner lore indicate that the shape of the constituent performances influences the shape of the respondent behaviors. For example, a game whose arrangement of constituent performances gives low power to an "It" person, will yield more frequent failures and desires to quit by the It than will a game giving high power to an It. Such a game will also produce informal anti-It alliances among non-It players. These respondent behaviors are not specified by the game's constituent performances but are made more likely by them (Gump and Sutton-Smith, 1955b). It is clear that other activities besides games have both the constituent and respondent behaviors.

The general hypothesis explored in the Gump and Redl investigations was that variations in the activity setting and in its pattern of constituent performances would be accompanied by changes in the respondent behavior of participants. As they are helpful in an effort to understand the classroom behavioral system, several studies from this investigation will be cited. The methodology in the studies was simple and direct. The behaviors of the participants—as those behaviors changed from one activity setting to another—were recorded in the specimen-record fashion described by Barker and Wright (1955). This method yielded a reasonably accurate picture of participant behavior and of the events that formed the context of the behavior.

In one study, the behavior of eight campers from three different cabin groups was recorded as these boys participated in swims, cook outs, and dining hall activity settings (Millen, 1958). The respondent behavior analyzed was hostile behavior. Table 1 shows the data. (Figures have been corrected for time variations.)

The frequency of hostile acts for each camper rose as he went from swim to cook out; usually it rose very sharply. For all boys except camper F, this frequency increased again as they went from cook out to dining hall. The significance level for variation among settings and across campers was .01; there was no significant variation among campers across settings.

TABLE 1

Number of Hostile Acts by the Same Boys in Three Camper Activity Settings

Camper	*Swim*	*Cook out*	*Dining hall*
A	0	11	58
B	0	34	54
C	6	18	24
D	10	11	97
E	9	26	28
F	12	78	25
G	7	15	32
H	13	33	64

SOURCE: Paul V. Gump and Jacob S. Kounin, "Issues Raised by Ecological and 'Classical' Research Efforts." *Merrill-Palmer Quarterly of Behavior and Development*, Vol. 6, 1959-1960, p. 146.

If one were to ask, "How much hostility will a boy at camp show?" the easy answer would be, "It depends upon who the camper is." These data do not support this answer. Instead, they suggest: "It depends upon which activity setting the camper enters." With only rare exceptions, the same counselors and peer associates were present in each setting; therefore, variations of leaders or child associates cannot have been responsible for variations in hostility expression. Apparently, there was something about swim that reduced hostility; something about cook out, and especially about dining hall, that increased it.

Some work has been done to identify what this something was; however, for purposes of explaining the general results, let us turn to one other population group in these settings, the counselors. Did their behavior vary from setting to setting? Although no records were taken focusing on counselors, they often interacted with boys who were observed; and their behavior appears in the boys' records. (Two counselors usually were present for each cabin-group activity.) A count was made of the number of times counselors interfered with, or intervened in, the behavior of campers. Table 2 gives the results.

TABLE 2

Number of Counselor Interventions in Three Cabin Groups in Two Camp Activity Settings

	Swim	*Cook out*
Cabin X	3.7	21.6
Cabin Y	2.2	12.2
Cabin Z	2.3	5.1
Totals	8.2	38.9

SOURCE: Paul Gump, Phil Schoggen, and Fritz Redl, "The Camp Milieu and Its Immediate Effects." *The Journal of Social Issues,* Vol. XIII, No. 1, 1957, p. 42.

Interventions by counselors in all three cabin groups rose as they went from the swim to the cook out settings (Gump, Schoggen, and Redl, 1957). Data from Table 1 suggest one reason why counselors became more interfering in cook outs—there was more trouble demanding such interference in cook outs. Inspection of the protocols revealed another possible reason for the difference. Because the two activities were different, the "built-in" tasks of the counselors were different. Constituent performances for a counselor in swim included keeping an eye on the boys, offering swim instructions if they were desired, and "being around." In cook out, the counselor's job shifted; here he became responsible for the meal. This meant seeing to it that a fire was built and protected, that food was protected and cooked, that boys did not burn themselves, and so forth. The activity involved task changes, and these changes made interference with boys' behavior more likely in cook outs than in swims.

One other study, involving boys in swim and crafts, yielded analogous results (Gump and Sutton-Smith, 1955a). First, social interaction among campers changed in amount and kind as boys and their counselors went from swim to crafts. Secondly, social interaction between campers and counselors also changed. For example, 30 percent of the interactions in crafts involved helping relationships in contrast with only 15 percent in swim.

Again, behavior of the adult manager of the children shifted because of the change in children's behavior and in counselor's tasks. In crafts, the constituent performances for campers involved procuring tools and supplies and then making something with them. Counselors were caretakers of tools and supplies and were superior in handicraft knowledge and skill. They were likely to be asked for help and likely to give help. This shift to helping was dependent upon the nature of the activity setting, not upon the training or disposition of children and counselors.

These studies, and others by Crawford (1957) and by Rausch *et al* (1959), bring home a general point: how participants behave is dependent upon the activity settings in which they are placed. Furthermore, the behavior influences reach beyond such obvious constituent behaviors as eating in an eating setting or playing in a game setting; the influence extends to social relationships among participants, to emotional responses of individuals, and to types of need gratification.

Concepts and Findings from Studies of Psychological Ecology

Barker and Wright (1955) have developed concepts and measurement methods for phenomena similar to activity settings. Their lable is "behavior setting"; and their conceptualization is both more broad and more systematically detailed than the one offered for activity settings.

Behavior settings involve all types of place-time-activity unities —not just child-centered settings. Examples of behavior settings are Sunday morning church service at the Centerville Presbyterian Church, activities at the corner drugstore, the high school Thanksgiving assembly, Mr. Jones's second-period algebra class, and home basketball game. Although the label attached to a behavior setting may refer to any combination of times, places, activities, persons, or things, it is understood that all of these factors contribute to the reality that is the behavior setting. A more complete label for the home basketball game is: "Basketball playing and watching by students and townspeople on some winter evening between 7:30 and 10:30 P.M. at the gym of Centerville High School." The behavior setting has two major aspects:

1. The nonpsychological milieu that is the time-place-activity pattern (7:30 to 10:30 weekday evenings, Centerville gym, balls, seats, baskets, and so forth).
2. The standing pattern of behavior (game playing by student athletes, watching and cheering by audience).

The nonpsychological milieu surrounds, supports, and limits the standing patterns of behavior.

The deceptively simple concept of behavior setting yields a unit of environmental analysis that has a number of virtues. The authors point out that behavior settings are visible, obvious, empirical facts about behavior; they are systems with internal and external dynamics; they are comprehensive (most of the observed behavior of persons occurs within behavior settings already identified); and they coerce behavior.

This point, the coerciveness of settings, is important and deserves further emphasis and attention. The changes in behavior of campers as they go from swim to crafts are settings-coerced changes. The change in behavior of school children as they leave spelling lesson and enter recess are settings-coerced changes. The influence of behavior settings is obvious when one reviews changes, concomitant with entering and leaving settings in his own behavior stream.

Sources of behavior-setting coercion rest in a number of areas. Barker and Wright (1955, pp. 55-57) discuss the possible sources of the "synomorphy" between behavior and the qualities of settings. Five of these sources relate to how settings limit or provoke behavior:

1. PHYSICAL SOURCES School hallways allow locomotion in certain directions. The absence of setting props (chairs or ledges) encourages standing and walking, discourages sitting or lying.
2. SOCIAL SOURCES Persons of power in settings influence others to conform to the appropriate behavior pattern. The teacher enforces a "stay in your seat" rule.
3. PHYSIOLOGICAL SOURCES Temporal and physical qualities of the setting affect behavior because of "built-in behavior mechanisms." For example, cold and heat affect behavioral tempo.

4. PHYSIOGNOMIC PERCEPTION The physical qualities of a setting seduce as well as compel. Many have observed that open spaces invite children to exuberant, romping behavior. Challenging or assertive behavior often results when some children occupy positions elevating them over other children. Social pressures from the en masse behavior of others also provide physiognomic directions. We tend to be sober in church but gay at a carnival.
5. LEARNING Over a period of time people learn the setting-required behavior. Once these behaviors are well-learned, a series of cues within the setting is sufficient to shape behavior; explicit pressures are not required.

The above list is a partial statement of the ways in which settings coerce behavior. Yet, the "why" of this is understood only dimly. The fact that coercion exists is apparent, and a few reasons for it are obvious. For our purposes, there remains the issue of the relevance of setting coercion to school and classroom behavior.

Setting Effects on Children's Behavior in School

An early work investigated the behavior of one second-grade boy, Raymond, in two classes of settings (Wright, *et al,* 1951). In one group of settings, Raymond was in a teacher-supervised classroom. In the other settings, he was in an adult-free situation. Certain aspects of Raymond's behavior and of his psychological habitat were measured by coding and quantifying material from a day-long specimen record.

Findings indicated that school settings (as compared with adult-free settings) yielded the following: less intense, energetic, creative, and satisfying behavior; more restless behavior (seven times more); smaller, less clear cognitive fields; less positive geniality; more social motivation; and fewer direct-interest motivations.

Although only one subject was involved in this study, results quantify some of the apprehensions many have held about the behavioral and experiential effects of classroom environments.

The study of Raymond suggests that psychological conditions for children in classrooms are not always intrinsically rewarding; however, another study indicates that the classroom does not produce intense or frequent manifestations of emotional disturbances in children. Through the use of specimen records of six different children, Fawl (1959) discovered that five of six children displayed lower disturbance frequencies in "school academic" than they did in the rest of the behavior settings in their day-long records. The intensity of disturbance in the classrooms was usually less than in nonclassroom settings. However, the durations of disturbances in the classroom usually were longer than in other settings. None of the significant tests yielded p's of less than .05. However, p values for frequency and duration were very close to the .05 level. Thus, classroom disturbances were not especially intense or frequent, but tended to show mild, prolonged resentment and boredom. Fawl reported (p. 106): "For the most part, the assignment of boring or unpleasant lessons accounted for the longer classroom disturbances."

Teachers' Behavior toward Children

The question of what happens in school settings was also investigated by Dyck (1963). How are children in school dealt with by teachers? How do teachers differ from parents in this regard? Among other things, Dyck asked:

1. Do teachers initiate social actions with the child more often than do parents?
2. Do reasons for the entry into the child's behavior differ between teachers and parents? (Dyck identified twelve different reasons for entry.)
3. To what extent do teachers' and parents' entries deal with the "requirednesses" of the setting? Dyck labeled these "requirednesses" *ritual-related.* Examples would be interactions about eating at home meals, about practicing one's arithmetic in school academic.

Five home situations in our culture that involve rituals were utilized—getting up, breakfast, lunch, supper, and going to bed. It was presumed that the forces acting upon parents to influence children to get on with the business of the setting would be greatest in these situations. Comparisons of teachers and parents (as their interactions with children appeared in specimen records) produced these sample results:

1. Teachers initiated 73 percent of all social contacts; parents initiated only 34 percent.
2. Teachers' reasons for entry centered around one single desire—to seek activity on the part of the child. Parents, on the other hand, entered more frequently for reasons dealing with the child's desire for activity. Parents were also much more responsive to the overt expressions and activities of the child.
3. Teachers' social interaction was far more related to ritual than was parents'.

The foregoing sample of Dyck's results supports the following conclusion regarding parents and teachers in those situations that press for adult shaping of children's behavior: Teachers, more often than parents, take the initiative in starting interaction, usually seek to move the child rather than to respond to his interest, and deal almost exclusively with the "business at hand." In brief, teachers dominate children more than parents.

While the study of Raymond dealt with only one child, the Dyck study involved twenty-one children in a number of ordinary classrooms. Results can be taken as indicative of the way teachers behave in large numbers of American classrooms. Other research shows that, in an English school, the dominance of teachers is even more extreme (Dyck, 1961b).

Dyck's reflections upon the reason for the more intrusive, influencing, business-centered behaviors of teachers are pertinent to the general orientation of this chapter regarding coercion of settings upon the adults responsible for children:

> The classroom is much more businesslike, much more demanding so that its agents and its subjects get on with the requirements of the setting. This is something most of us more or less take for granted. In Midwest homes, even in situations which made definite demands upon parental behavior and, in turn, upon children, there was time to permit social interaction about matters unrelated to the ongoing ritual and mechanics of the settings. Home, in the Midwest, at least, did not have any settings that were as inflexible in prodding the child's behavior as the academic classroom. (Dyck, 1963, p. 97.)

Educationists devoted to the democratic philosophy have urged that teacher-pupil interaction be less characterized by teacher domination. However, make no naive expectations regarding the quality of interactions feasible in the usual academic classrooms.

The nature of the settings themselves may coerce social interaction patterns as much as teacher personality variables.

A study by Hughes *et al* (1959) is relevant to this issue of setting-coerced teacher behavior. Using an adaptation of specimen-record methodology, Hughes took specimen-record samples of thirty-five elementary school teachers' behavior in various teaching situations. Thirty of the thirty-five teachers were rated "good" by personnel who had opportunity to know their work. Only five teachers were rated as "causing concern" to school people. The specimen records were content analyzed into the following seven large categories of teacher social activity: controlling functions, impositions of teacher, negative affectivity, facilitating functions, functions that develop content, functions of personal response, and functions of positive affectivity.

The functions of controlling, imposition, and negative affectivity were combined by Hughes into a dominative-behavior index (Hughes, *et al,* 1959, p. 293). The range of dominative behavior by the teachers was from 48 to 75 percent with a median of 59 percent. *No* teacher—from a group of thirty-five predominantly "good" teachers, from five different school districts—was able to hold his dominative actions at significantly less than one half of his total social output.

Data from the Dyck and Hughes study point up the predominantly controlling aspect of the teachers' behavior in ordinary American classrooms. Differences in degree among teachers and effects of those differences have been found, but the general tenor of the findings is that "something" is holding up the general level of teacher dominance in social relationships with children. In spite of decades of pressure and persuasion from educational philosophy and mental hygiene circles, this "something" still operates.

What Determines Teachers' Social Behavior toward Children?

It has been reasonably well established that some of the differences in behavior and effectiveness of teachers in relation to children is correlated to differences among teachers as persons.

(Heil, *et al,* 1960, is one of the more sophisticated examples.) In contrast, the Hughes data indicate that teachers' social behavior also varies with teaching situations.

Hughes' observations were taken in several different curricular situations—reading, social studies, arithmetic, and activity period. The total number of social acts and the number in each of the seven categories were tested for differences among teachers and among situations. In all but one of these comparisons, situations showed statistically significant differences. In general, the activity period was the most variable. In contrast with social studies and reading periods, activity period produced: (1) fewer total acts, (2) less controlling, (3) less content development, (4) somewhat less positive affectivity ($p = .06$), (5) more personal response, and (6) more negative affectivity.

Variations among teachers across situations presented two fewer significant comparisons than did variations of situations across curricular situations.

The author appears to attach only secondary significance to these findings; she quite correctly points out the necessity for observation of a teacher in various situations in order to achieve a representative sample of his behavior. *But suppose one stops this quest for describing and categorizing teachers?* Suppose one wishes to know the factors that produce teachers' behaviors? In that event, data such as these should be taken quite seriously. They indicate, as did the data on counselors in camp situations, that the context shapes the behavior of adult managers. Furthermore, this shaping is somewhat independent of variations in personality. We tend to dismiss such data because it seems so obvious. Once one thinks of it, it seems quite apparent that an activity period is likely to involve (1) more independent work by students (resulting in fewer teacher acts and less controlling), (2) less academic effort by teachers (resulting in less content development), and (3) more shuffling about and opportunity for deviancy (resulting in more negative affectivity).

Behavior-Setting Management and Teacher Effectiveness

It is proposed that the implications of findings such as those in the Hughes study be pushed as far as possible. It is proposed, further, that research in teacher effectiveness be directed, by some researchers at least, toward that aspect of teaching which might be called *behavior-setting management.*

The necessity for maintaining optimum learning environments is not a new idea in educational thinking. Hughes refers to the idea a number of times. MacDonald explicitly states the proposition that "the task of the teacher is to organize and arrange the learning environment" (1960, p. 24). In spite of references to the learning environment, research in teacher effectiveness has not dealt with environmental manipulation in any comprehensive fashion. One substitute for this approach has been an emphasis on the traits and personal-social qualities of teachers. There can be no doubt that this line of investigation can demonstrate significant relationships between teacher qualities and pupil-response variables. Evidence on this point is cited elsewhere in this book (Chapter 3). However, this approach seems inadequate for several reasons.

First, teachers' behavior toward children is influenced by the position teachers must take in a given activity structure. Why do counselors offer more help to campers in the craft shop than in swimming? It is clear that, as the protector and dispenser of materials (and as the activity-provided consultant), the counselor is required to provide such help in crafts. Why do teachers in activity periods develop less content than do teachers in social studies classes? Obviously, because the *role* of the teacher in activity period is one of helping children deal with contents already developed. Once an activity is underway, the *role* of a teacher is not a matter of teacher disposition or personality. The teacher will have to engage in a pattern of behavior consonant with his position if he is to be effective.

Second, the teachers' behavior with children is influenced by the behavioral feedback of these children as they deal with the particular supports and coercions of the activity structure. Studies have indicated a correlation between negative aspects of teachers' social behaviors and negative aspects of pupil behaviors. However, some of the worst teaching occurs when the teacher faces apathetic or aggressive pupils. Forces producing teacher anger, confusion, or domination grow out of the harassing quality of the teacher's contemporary psychological situation. Logically, we could ask what the class does to the teacher as readily as we might ask what the teacher does to the class. Why do counselors interfere so often in cook outs and so rarely in swims? The answer: the activity structure in swims is such that campers are invited to perform satisfying behaviors that do not require much interference. In cook outs, the focus on gratification possibilities (fire and food) invites boys to engage in dangerous or "illegal" acts; the result is required police action on the part of counselors.

Third, description of teacher social action is inadequate in a description of classroom environment. It is important to understand clearly the locus of influence behind factors generated by the activity setting. Teacher behavior probably *is* the most salient factor in the classroom situation. However, teacher behavior has many aspects; only certain of these are represented directly in analysis and description of personal-social qualities exhibited during interaction with children. The teacher can arrange time, space, objects, pupils, and himself into various constellations of learning environments. His social behavior in class involves, in part, the manner in which he does this, and, in part, a reflection of the possibilities and necessities imposed by the arrangement itself.

Classroom Saliencies as Perceived by Students

The fact that teachers' personal-social behavior is not the only prominent aspect of the classroom environment is illustrated by data from an exploratory study now in progress at the Midwest Psychological Field Station. Subjects were 196 junior-year students in five different high schools. Instead of requiring students to talk about their teachers, the investigators invited students to tell about their classes. These students were asked to identify as "plus" the several classes "which were especially good, satisfying, worthwhile." They were then asked to explain for each: "what I did, what went on, what made classes good, 'plus'; what I enjoyed or got out of the classes." Similar questions were used for "minus" classes. This open method invited the subject to reveal the salient aspects of classroom situations as experienced by him.

Student responses fall into five major categories: the teacher (his qualities and behaviors), the subject matter, the classroom activity, the classmates (their qualities and behaviors), and the working conditions. Results are presented in Table 3.

TABLE 3

Distribution of Major Locations for Plus and Minus Reactions to High School Classes

	Teacher	*Subject matter*	*Activity*	*Classmates*	*Work conditions*
Number of Times Category Used	267	506	248	37	17
Percent of Times Category Used	25	47	23	3.5	1.5

Subject-matter considerations most frequently demonstrated satisfaction and dissatisfaction in classes. Next were teacher qualities and activity qualities (to equal extent). Evidently, high school juniors do attribute their plus and minus experiences to more than the personal and social variables of their teachers.

The locations given in Table 3 are gross; but it was possible to subdivide the teacher, subject-matter, and activity areas so as to yield a finer code of twelve possible locations. The categories of this code are listed below in order of their prominence.

These rankings, again, indicate the relative importance of variables of class environment that are not centered on the teacher's social behavior. This is not to say that the teacher is not responsible for some of the locations; in some cases he undoubtedly is. However, students do not see the teacher's personal and social qualities as the most prominent factors in their classroom environments.

TABLE 4

Distribution of Teacher, Subject-Matter, and Activity Areas into Twelve Locations

	Percent
Valence toward Subject Matter	27
Valence toward Activity	20
Assessment of Teacher's Pedagogical Ability (Answers Questions, Interesting Talks, and So Forth	13
Utility of Subject Matter	12
Reaction toward Teacher's Personal and Social Qualities (Friendly, Nice, and So Forth)	9
Ease or Difficulty of Subject Matter	8
Classmates (Fun To Be With, Unruly, and So Forth)	3
Utility of Activity	2
Teacher as Disciplinarian	2
Working Conditions	1.5
Ease or Difficulty of Activity	1
Teacher as Vaguely "Good"—"Not Good"	1

Summary

Data and discussion thus far lead to three generalizations.

1. Interdependency of many classroom elements, as illustrated by the study of ripple effect, is a fact, not a theoretical assumption. This interdependency of multiple parts implies that utilization of system variables, as opposed to controlled single variables, may provide a feasible and fruitful approach to the issue of teacher effectiveness in guiding classroom events.
2. A system variable coercing both the behavior of children and the behavior of their adult managers is referred to as the activity setting or behavior setting. The assertion that settings are coercive seems to be a restatement of the well-accepted formulization that behavior is a function of environmental and personality forces. However, researches in education rarely have employed settings as independent variables. Instead, the approach has

been to employ restricted variables such as teachers' personal and social qualities.

3. Since settings coerce both pupils and teachers, it follows that teacher effectiveness can be improved if this coercion is properly understood. Research efforts to develop an understanding of settings will require concepts that identify setting units and give dimensions to their qualities. Once this is done, the relationships between setting variations and individual student behaviors (including learning behaviors) can be delineated.

The remaining discussion considers possibilities in conceptualization of the classroom environment and its settings.

Classrooms as Learning Environments

Problems in Conceptualization of Environments

Efforts to describe environments are not new in psychology. The problem has been to create comprehensive and behaviorally relevant descriptions. Descriptions of physical environmental factors are objective, but these descriptions omit factors that have considerable impact on behavior. Sherif and Sherif (1956, p. 121) outlined an objective—yet behaviorally pertinent—approach to the description of social situations. They classified situational factors into areas related to (1) qualities of the individuals behaving in the environment, (2) the task or problem faced by the individuals, (3) the site and facilities involved, and (4) relations between individuals, problems, and facilities. For some purposes, certain of these factors may or may not be considered to be part of the environment. For example, the relationship actually established by individuals (given certain task and role requirements inherent in the environment) might be considered a variable dependent on these requirements.

A discussion by Sells (1959), as well as observations at a recent seminar on the taxonomies of situations (Symposium, 1961), also are relevant sources for concepts in environmental theory.

A recent attempt to separate logically environmental phenomena from the behavior of individuals within environments has been made by Barker (1960). Briefly, according to Barker, the clue to the specification of the relationship between individual behavior and environment resides in distinguishing systems that obey different sets of laws but that involve influence of one upon the other.

Classroom Environments as Behavior Settings

The general proposal offered here is that classrooms offer environments that may be described as behavior settings.[3] This proposal is made for the following reasons:

1. Promising methodological and conceptual tools for the identification and description of behavior settings already have been devised (Barker and Wright, 1955).
2. The above are sufficiently comprehensive to include the behaviorally relevant factors necessary for valid prediction, from environments, of behaviors of individual participants.
3. The coerciveness of behavior settings upon individual behavior of both leaders and participants already has been established.
4. As the responsible leader in behavior settings, the teacher is in a position to manipulate these settings to fit his learning strategies.

A primary problem in conceptualization of environments involves a determination of environmental units. Since the environment presents a complex array of possible areas, it is important to decide upon the limits and size (inclusiveness) of areas to be investigated. One method for distinguishing among related areas has been offered by Lewin (1951). This method employs a degree-of-independence index. Independence degree refers to the extent to which changes in one area are reflected by changes in another area. If changes in area A are accompanied by a similar amount of change in area B, A and B are said to be parts of one unit; they show high interdependence or low independence. Obviously, whether A and B are separate units or parts of one unit will depend upon the level (or cutting point) of independence established. A high independence cutting point results in more inclusive units and fewer units per array of areas. Low independence points yield reverse results.

---

[3] Waimon (1961) has reported promising use for the behavior-setting concept for organizing classroom observational methods.

Barker and Wright have used an adaptation of this to delineate the environmental areas they call behavior settings. They reasoned that independence among related areas could be estimated by judging the extent to which elements in each area were different. In order to determine if two related portions of the environment represented subareas of one behavior setting or two discrete settings, they rated the extent to which elements within them were discrete. The elements employed were leaders, participants, times, places, behavior objects, molar action, and behavior mechanisms. A theoretical range of independence from seven to forty-nine was derived. Areas with an independence value of twenty-one or more were considered separate settings. Barker and Wright refer to this measure of estimated independence as the K-21 index.

An analysis of a school's total environment demonstrates that behavior settings of the K-21 size level have good utility. A child's day, or year, in school is spent in a number of behavior settings—in academic classrooms, music periods, art classes, hallways, offices, playgrounds, club meetings, and so forth. One way to describe the child's environment is to describe the behavior settings he enters and the positions he occupies within these settings. Student-life investigations employing this method already are in progress at the Midwest Psychological Field Station. One complete study deals with the relationship between school size and amount and degree of responsibility of student participation in extra academic behavior settings (Barker and Gump, 1964).

However, when interest is limited to environmental units *within* a classroom, too few units are yielded by the independence cutting point of K-21. For example, in one elementary school the behavior setting "third-grade academic" represented all of the academic environmental occurrences for that particular teacher and class for an entire year (Barker and Wright, 1954, p. 156). If one desires units within the classroom it would seem that a lower independence cutting point would be required—areas within the classroom could then become subsetting units.

These subunits are perceivable by anyone who observes the flow of events in a classroom. An arithmetic lesson may start with a review discussion period. There follows a presentation of new concepts by the teacher. Finally, there is a period of seat-work exercises for drill in the new concepts. Although the divisions employed for classifying "within classroom" events usually have been based upon subject-matter distinctions, this method is not necessarily the most effective. In the arithmetic lesson just described, the marked differences in subsettings would have been ignored by a subject-matter division; the whole environment would have been called "arithmetic." If the method employed by Barker and Wright were utilized, divisions in the classroom would be based on the degree to which parts of this setting displayed different times, places, leaders, participants, molar actions, and behavior mechanisms.

It is clear that there are greater independence ranges for some of these criteria elements than there are for others. The leader tends to be the same person throughout various subsettings (since the leader is the teacher). Participants vary, however, as when certain subwork groups are formed. There also are time and place variations in parts of the classroom. Finally, there are differences in behavior objects and mechanisms and in molar behaviors.

In summary, units within an environment may be established by use of an independence index. The cutting point used by Barker and Wright serves well to establish larger units (behavior settings) within a school. The cutting point used for independence of behavior settings is too high for use within classrooms because often it yields only one unit. It would appear feasible to use the Barker and Wright method with the establishment of a lower cutting point. In this way, the unit of analysis would be consistent with the large behavior-setting unit used to describe total school environments.

A second concept from Barker and Wright's analysis of environments is *variety*. The independence or K-21 index is useful in distinguishing among discrete settings. The variety index refers to differences in kind. Thus, two grocery stores represent two quite discrete behavior settings—but are of the same kind; a grocery store and a swimming pool are discrete settings of different kinds. The variety classes of settings vary to the extent that they employ dissimilar action patterns, behavior mechanisms, loci, and behavior objects. The number of behavior-setting varieties entered by an individual is indicative of the variety of experience opportunities offered by that individual's environment. This is true because, by definition, behavior settings of different varieties induce or support different perceptions and behaviors.

The cutting point now used for distinctions between settings on the variety index is too high to reveal differences within the classroom. It would be possible to lower this cutting point to obtain a variety measure. It would seem that the problem of development of a variety index for "within-classroom" subsettings is feasible.

Such an index for classroom activity might serve a very useful purpose. One of the complaints sometimes made about students is that they are apathetic; it is quite possible that lack of variety contributes to this apathy. Environments that require the same molar behaviors to be repeated—with the same mechanisms, and objects, and in the same sites—are likely to yield apathetic responses.

To date, investigations of the effects of subsettings within classroom environments are few. We are indebted to Kowatrakul (1959) for one example of the way in which such studies might proceed and what kinds of yields might be expected. The independent variables in the Kowatrakul study were subject-matter areas and classroom activities. Subject areas were: science, language, social studies, and arithmetic. Classroom activities were: independent seat work, discussion, and watching and listening. It is essential to recognize that activities were not identified by reference to the responses of individual pupils; rather, they were action descriptions of the subsettings within which individual students behaved. The activity labels were brief statements of what Barker and Wright would describe as the "standing patterns of behavior" for classroom subsettings.

The dependent variable in the Kowatrakul investigation was the individual behavior of students when they were within the subject areas and the activities (or subsettings). Behavior categories were: intent on work (on-going), social-work oriented, "social-friendly" (nonwork oriented interaction), withdraws momentarily, intent on work in another (not current) academic area, and "work" in nonacademic area (for example, making a paper airplane or a chain of paper clips).

Kowatrakul investigated two classrooms, one in the fifth grade and the other in the sixth. She applied analysis of variance to her data and demonstrated a number of significant relationships between student behaviors and variations in subject matter and activity. In spite of the fact that teachers, pupils, and physical conditions were different in the two grades, certain relationships between subsettings and student response were found in both classes. For example, student behaviors of the "illegal" type (social-friendly and "work" in nonacademic area) consistently and significantly were most frequent in the discussion subsettings. Independent work, on the other hand, was most frequent in seat work and in watching and listening. The Kowatrakul study demonstrated that activity or subsetting variation can account for the extent to which students apply themselves to the assigned work. It should be noted that variations in teacher personality were not an issue; these were held constant in each class. Variations in environments managed by teachers produced the obtained results. One practical implication of Kowatrakul's findings is that the discussion subsetting is relatively difficult to handle and probably requires extra skill and planning.

A result not stressed in the research report but apparent from the data was that activity variation usually produced more widely discrepant student response than did variation of subject matter. For example, it was possible to compare the range of means in five of the dependent variable categories when these ranges were associated with subject-matter variation and when they were associated with activity variation. Since two classes were involved, ten measures for range were available. Subject-matter variation produced an average range of student behavior means of 3.26 percentage points; activity of subsetting variation yielded a range of 10.51 points.

A clear inference from this result is that student responses to areas of classroom events are related more to activity-pattern (subsetting) variations than they are to subject-matter variations. It might also follow that differences in teacher behavior (such as those revealed in the Hughes study cited above) would be sharpened if activity-pattern distinctions were substituted for subject-matter distinctions as employed by Hughes.

Discussion of classroom environments thus far has centered upon the identification of classroom subsettings. A possible method of measuring variety among subsettings and a description of one study illustrating the coercivity of subsettings was given. There remains the question of concepts for the description of subsettings once they are identified.

Learning Formats for Classroom Subsettings

Part of the behavior in a particular subsetting reveals an understood activity pattern, a kind of "the way it is supposed to go" idea, shared by teachers and pupils, which partially accounts for the maintenance and progress of the observed subsetting. This activity pattern is a cluster of constituent performances similar in principle to those mentioned earlier in reference to children's games. Learning is the primary consideration in the classroom, and the pattern of performance is designed to facilitate learning. Thus, the label "learning format" might be applied to that aspect of the subsetting dealing with its pattern of constituent performances. The learning format specifications can relate to any and all of the following activity issues: *who* shall do *what,* with *which* objects, *when, where,* and with *whom.* The relevance of the learning format to the problem of teacher effectiveness lies in the fact that teachers can manage formats and, thereby, influence the quality of behavior settings and subbehavior settings surrounding their teaching activities.

Although the format is *reflected* in behavior, it is *located* in the conceptions of the classroom participants. The teacher has an idea of the pattern of constituent performances (with associated objects, times, and places) that he would like to see translated into a behavioral reality.[4] Eventually, this idea of the format will be shared, to some degree, by students. The students' conception of the format develops from a number of sources. The teacher may inform students; workbooks and chapter endings define format behaviors; the arrangement of physical objects and spatial and temporal differentiation suggest which behaviors are appropriate; or, for any one student, the behavior of other students can serve as a format communication. However the format may be developed and communicated, the working classroom group operates with an understanding of how the business of learning is supposed to proceed.

---

[4] The teacher may not be the only holder and proposer of format ideas. It is possible for students to conceive and suggest formats. The teacher's conception of format is emphasized because this simplifies explanation and because the author believes that the teacher most often takes responsibility for format conception and specification.

In order to illustrate the possibilities for description of learning formats, it may be helpful to sketch four somewhat old-fashioned methods for learning a list of spelling words. To facilitate reference to these formats in subsequent discussion, each will be given a brief label.

COPY CHECK Students copy a word in the list twice; then they cover the copied words and write the word from memory. They then uncover the copied words to check correctness. Each word in the list is handled in the same way.

PAIRED TUTORS Students are separated into pairs. One student pronounces and spells the word, the other student repeats the word and its spelling; the first student checks accuracy. After the list is finished, the student roles are reversed.

TEACHER-LED CHANT Students look at a blackboard on which each word is printed in large letters. The teacher points, in succession, to each letter of a word and asks the class to chant the letters. Students chant again with the word covered. The teacher listens to see if they are spelling correctly.

OPEN ASSIGNMENT Students are given a list on Monday and are told to learn the words it contains before test time on Friday.

It is possible to develop dimensions that describe variations among these simple formats:

1. MEANS PRESCRIPTIVENESS extent to which the ways of participating or learning are defined by the format. High means

prescriptiveness is apparent in all of the formats except open assignment. One might consider that this latter method is not a format at all except for the fact that *what* is to be learned, *when* it is to be learned, and *who* is to learn are all specified.

2. INTRUSIVENESS OF TEACHER ROLE extent to which teacher's behavior (once the format is in operation) is required in order to continue the format. Teacher intrusiveness is very low in copy check, paired tutors, and open assignment; it is very high in teacher-led chant. Note that this dimension does not involve everything the teacher might do during this subsection of class work; it refers only to format-prescribed actions. It easily can happen that the feedback from excessive student noise in paired tutors will require a policing behavior—but this action is not format specified.

3. NECESSITY OF PUPIL SOCIAL INTERACTION extent to which the format sets up interlocking behavior patterns for pupils. Copy check is very low; any social interactions in this context would occur in spite of the format. Paired tutors is very high. Teacher-led chant calls for a kind of togetherness. Open assignment cannot be rated because of its low prescriptiveness in this area.

4. FLEXIBILITY OF PROGRESS extent to which students can deal, at their own pace, with learning steps. Copy check and open assignment are high in such flexibility; paired tutors and teacher-led chant are low.

5. USE OF ACTIVE BEHAVIOR MECHANISMS degree to which active, "doing," body actions are involved—as opposed to passive or covert mechanisms. Copy check is relatively low, involving much looking, thinking, and restricted manipulation. Paired tutor and teacher-led chant involve talking (verbalizing) as well as looking, listening, and thinking; these two formats seem higher on this action dimension. Open assignment cannot be rated because the actions to be taken by the students in order to learn are not prescribed.

The above dimensions are meant to serve as samples of concepts that could be used to describe formats. It is possible to define additional dimensions or to construct subdimensions within those given above. Intrusiveness of Teacher Role can be subdivided into intrusiveness of various roles—teacher as content giver, monitor, evaluator, opener and closer of action units, and so forth. Particular dimensions can be chosen according to the kind of independent and dependent variables of interest and according to the extent to which a dimensional point is actually specified by the format.

A learning format with very high prescriptiveness is the teaching machine. The format of algebra study on teaching machines precisely and emphatically specifies a number of activity factors. Intrusiveness of teacher role is very low because the machine, not the teacher, is content giver and evaluator. Necessity of pupil social interaction is very low. Flexibility of progress is high. The format includes behavior mechanisms of looking, thinking, and manipulation; it excludes talking, listening, and gross motor activity. Other readily apparent dimensions are immediacy of feedback and continuity of overtly active participation; the teaching machine format is high on both dimensions.

Types of Variables Dependent on Format Variation

It would be agreed that the ultimate dependent variable in the classroom is learning by pupils. There is, however, some distance between the format understandings of teacher and pupils and the learning variable.

Some areas of teacher, group, and individual behavior are not prescribed by a format but are relevant to one. For example, a format of low means prescriptiveness may have a number of *likely*, but not *specified*, behaviors. In a format of low means prescriptiveness, students may make heavy demands upon the teacher for advice and technical help. High group interaction may result from the necessity to get direction and confirmation from "somebody." Because such a format does not tell pupils exactly what to do, individual students may show high creativity in filling in gaps left by the format, or, on the other hand, they may desert the format completely. If formats continually exclude the use of active behavior mechanisms that children's bodies require (such as talking, gross motor locomotion, and manipulation), the result may be individual restlessness and task aversion. Then, subgroups may begin nonformat interactions. Finally, the teacher may be pushed into prodding, policing action. Neither student restlessness nor teacher policing are required by the format, but both are made more likely by it.

The above examples point out a need to establish the sort of student variables that must be taken into account when format choices are made. Pupils' responses to a given format will depend upon (a) their abilities, (b) their interests, and (c) their tolerances. Sophisticated choice of format probably requires not only selection according to relatively enduring traits of the pupils in the group, but also a choice that takes account of current group moods. One issue in format development, then, is "what kinds of formats for what types of individuals and groups?"

A special set of behaviors dependent upon format are those related to learning. Different formats, under given conditions, will lead to different amounts of student responses that are learning-oriented. For example, the discussion format in the Kowatrakul study yielded fewer learning-oriented student behaviors; the independent seat-work format resulted in more such behaviors.

More basic than the issue of the frequency of learning-oriented responses is the issue of the amount of actual learning achieved. Eventually research must discover the capacity of various formats to increase student learning. This question is complicated because a particular format may advance some learning purposes but not others. A teaching machine might produce content mastery but teach nothing about how to be a contributing member of a task group. The fact that American educational goals are varied increases the need for research and the development of concepts in the area of "which formats for which learning goals."

Concluding Thoughts

In this chapter the purpose has been to call attention to a new orientation to the problem of teacher effectiveness. In brief, this orientation de-emphasizes the teacher's personal qualities and social relationships and focuses instead on the teacher's ability to develop a learning environment.

Teachers may be wiser in their choices of relevant variables than are many educational researchers. Teachers know that their job is to manage learning activities; they realize that this is their first, their most obvious, and their most necessary function. Teachers know that they work through the mediums of activity and props, that format variation is one of the strategic measures available to them for purposes of influencing the quality of their classroom behavior settings. Most of the variations in this area have been left to the inventiveness and good sense of practitioners. The dramatic exception is the teaching machine, upon which much scientific talent has been lavished. Is it not possible that other learning formats could be researched to accomplish other educational goals? Teacher effectiveness might be significantly increased by scientific attention to the ecology of learning.

7

NED A. FLANDERS

Some Relationships among Teacher Influence, Pupil Attitudes, and Achievement

This chapter is divided into two parts. The first is concerned with a research program—conducted at the University of Minnesota and supported by the U.S. Office of Education—in which an analysis was made of teacher statements, pupil attitudes, and achievement. At appropriate points, the findings of other research studies are cited to support or modify the evidence from Minnesota. The second part of the chapter is concerned with the implications of this research for the practical aspects of evaluating teaching.

Research on Teacher Influence, Pupil Attitudes, and Achievement

Teaching may be conceived of as a series of overt acts over a period of time. How many different kinds of acts to be seen depends upon what one is looking for. Suppose one restricted himself to verbal behavior and chose to classify communication into three categories; for example, teacher talk, student talk, and silence or confusion. For only three categories based on who was talking, observer reliability would be high, but the psychological usefulness of the data would be limited. Suppose one had a great number of categories based upon such microscopic detail that each verbal act was seen as slightly different from the next. With so much to keep track of, observer reliability would be very low or nonexistent; the efficiency of the observation process would be very low and tabulation problems high; and the data would be too complex to analyze.

Somewhere between too many and too few categories is an optimum number of categories that will permit the observer to reach psychologically useful inferences by recording acts of influence as rapidly as they can be identified. The limitations of such a system will arise from the nature of the categories and the skill of the observer—compared with the purposes of the observation.

In the research project to be described, ten categories were used to classify the statements of the pupils and the teacher at a rate of approximately once every three seconds. It was found that an observer can be trained to categorize at this rate with sufficient accuracy (Flanders, 1960b) and that, in a very short time, the teacher and students can adjust to the presence of the observer.

The ten categories included seven assigned to teacher talk, two to student talk, and one to silence or confusion. When the teacher was talking, the observer decided if the statement was: (1) accepting student feelings; (2) giving praise; (3) accepting, clarifying, or making use of a student's ideas; (4) asking a question; (5) lecturing, giving facts or opinions; (6) giving directions; or (7) giving criticism. When a student was talking, the observer classified what was said into one of two categories: (8) student response or (9) student initiation. Silence and confusion were assigned to category (10).

In practice, an observer also kept a record of the different periods of classroom activity. In an hour he collected observation data separately for all periods of routine administration, such as settling down to work, cleaning up, passing out materials, and collecting materials; for all periods of evaluation in which homework or tests were being corrected; for all periods in which new material was being introduced; for all periods in which the students and teacher planned activities together; for all other class discussions; and for all periods in which the students worked at their desks or in groups. Thus, at the end of an hour's observation, it was possible to add up the different kinds of statements for each of these six types of classroom activity separately and to combine these into a grand total for the entire hour.

This method of observation is called "interaction analysis" in the classroom, and it can be used to quantify the qualitative aspects of verbal communication. The entire process becomes a measure of teacher influence because it makes the assumption that most teacher influence is expressed through verbal statements and that most nonverbal influence is positively correlated with the verbal. Those who have worked with this technique are disposed to accept this assumption.

Interaction analysis is a specialized research procedure that provides information about only a few of the many aspects of teaching. It is an analysis of spontaneous communication between individuals, and it is of no value if no one is talking, if one person talks continuously, or if one person reads from a book or report. Unless additional records are kept, the following kinds of information will be ignored—right, wrong, good, or bad content information—whatever is being discussed; the variety of instructional materials being used; the various class formations during learning activities; the preparation of the teacher as revealed by lesson plans; and anything else not directly revealed by verbal communication. Of the total complex called "teaching," interaction analysis applies only to the content-free characteristics of verbal communication.

The Point of Departure

The early research designs were simple. It was merely assumed that, by identifying classrooms in which the students had constructive attitudes and those in which attitudes were less constructive, comparisons of the contrasting teacher influence patterns would reveal significant differences. This design worked well from 1954 to 1957, the early phases of the research program.[1]

The student attitudes chosen for measurement are illustrated by the attitude inventory used in a 1957 study of New Zealand elementary-school classrooms.[2] The five scales used were: teacher attractiveness—liking the teacher; motivation—finding school work interesting; rewards and punishments—feeling that rewards and punishments were administered fairly; independence—feeling free to make some important decisions and to direct oneself while at work; and disabling anxiety—certain paranoid reactions to the teacher's authority.

After the New Zealand study the next logical step seemed to be an approach to the question of whether or not the constructive attitudes of the students would be positively correlated with measures of achievement. A contract with the Cooperative Research Program, U.S. Office of Education, 1958-1960, provided this opportunity.

First, sixteen teachers were selected from a population of seventh-grade combined English-social studies classes. This, because test evidence was available indicating that the classes chosen would represent the wide range of teacher-influence patterns of the parent population. A similar sample of sixteen eighth-grade mathematics classes was selected by the same methods. All of the instructional materials needed for a two-week unit of study were developed, and they included such a variety that each teacher could use his natural style of teaching. The teachers had adequate time to look over the materials and plan the unit before it was scheduled in one of their regular public school classes. No teacher saw the carefully prepared achievement tests before teaching. However, they did know the objectives of the tests and that the tests would be designed to measure not only the knowledge and skills of problem solving but also the application of knowledge and skills to unique and unusual problems. The tests were administered before and after each unit so that adjustments for initial ability could be made. In addition to the achievement test, the members of the research staff also administered an attitude inventory based on the earlier research instruments.

---

[1] During the early period the research program was sponsored by the Laboratory for Research in Social Relations, University of Minnesota, a department of the College of Science, Literature, and the Arts.

[2] The New Zealand study was made possible by a Fulbright Research Scholarship and the cooperation of the New Zealand Council for Educational Research, Mr. George Parkyn, Director. Results of the New Zealand study are reported in Flanders (1959).

During 1959-1962 a project concerned with the application of interaction analysis to the in-service training of teachers was completed with the support of a grant from the Educational Media Branch, U.S. Office of Education. Five sound filmstrips were developed to help teachers understand the purposes and limitations of interaction analysis, to use it as a tool for analyzing their own verbal behavior and the verbal behavior of their colleagues, and to see if carefully designed in-service training programs will result in actual changes in the spontaneous patterns of teacher influence. Heretofore, the appraisal of in-service training has been informal or, at best, an analysis of possible changes in the attitudes of the participating teachers. Interaction analysis provides a tool to find out if changes in attitude result in differences in teaching behavior.

Theoretical Principles of Teacher Influence and Early Research

The research program, up to 1958, was designed to find out if certain patterns of teacher influence could be associated with pupil attitude scores. When a distribution of verbal statements within the categories is plotted separately for classrooms in which scores on constructive pupil attitudes are low—in contrast with classrooms in which the same scores are high—a histogram (see Fig. 1) results.

The two profiles in Figure 1 show the proportion of different statements for two imaginary composite teachers. The percentage falling into each category is based on all the statements of five teachers whose classes scored highest and five whose classes scored lowest in a sample of thirty-three teachers of standard four, in New Zealand. The data from earlier studies of elementary and secondary classrooms in Minnesota were similar but not identical.

Conclusions drawn from these earlier studies are consistent with each other in spite of differences in grade level and the fact that differences exist between New Zealand and Minnesota schools (Flanders, 1959).

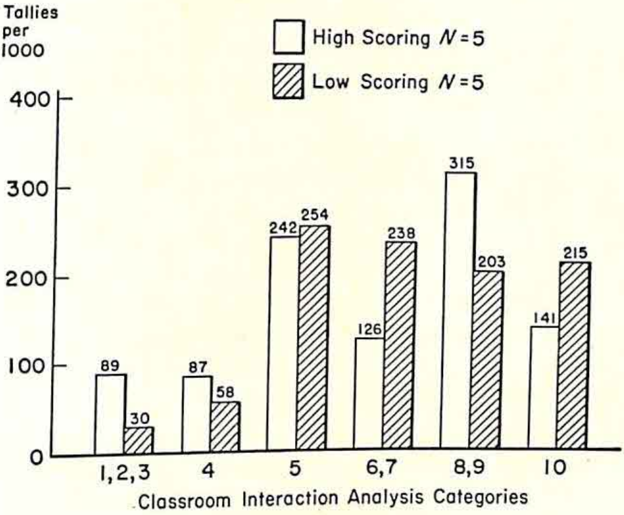


Fig. 1. Interaction-analysis data from New Zealand standard-four classes scoring high and low on the student attitude inventory.

All classrooms visited had in common a situation in which the teacher was clearly in charge, using his authority to control classroom activities. There was the impression that the teacher could do very little to avoid controlling classroom activities. When a teacher did succeed in withholding control or influence, there were predictable outcomes. The authority of the teacher was not only a psychological fact, it was the basis of school law. Later on, differences in the way teachers use their authority will be discussed at length. This will include discussion of the way in which some move in to control directly and others provide support to students who seem to control themselves.

Another element common to all classrooms was emphasis on subject matter. In both Minnesota and New Zealand, the knowledge, skills, and attitudes associated with content were given primary emphasis. The particular categories used for observation tended to ignore this primary emphasis and to concentrate, instead, on the management skills of the teacher. The way the teacher encouraged or restricted student participation was isolated for special attention by the very nature of the categories. The fact that these management skills can be shown to be related to subject-matter achievement indicates their importance to classroom learning even though they are not emphasized in classroom discourse.

The differences noted between the high and low scoring classrooms in Minnesota and New Zealand made possible the development of a set of hypotheses about teacher influence and student attitudes related to learning.

> ***Hypothesis 1*** Restricting student freedom of participation early in the cycle of classroom learning activities increases dependence and decreases achievement.
>
> ***Hypothesis 2*** Restricting student freedom of participation later in the cycle of classroom learning activities does not increase dependence but does increase achievement.
>
> ***Hypothesis 3*** Expanding student freedom of participation early in the cycle of classroom learning activities decreases dependence and increases achievement.

Each concept used in the three hypotheses will be given an explicit theoretical meaning and the steps necessary to measure each will be outlined briefly.

Certain teacher statements inhibit student participation and others encourage it. Restrictions occur when a teacher lectures or expresses his own opinions (category 5), gives directions (category 6), gives criticism or justifies the use of his own authority (category 7). He uses these statements to exert direct control. He is directive in order to focus on a particular idea or problem, in order to achieve compliance, or in order to correct misbehavior. In doing this, he will talk more and will take a more dominant role in the classroom.

Expansion occurs when a teacher asks questions (category 4), clarifies student ideas and uses them in problem solving (category 3), praises and encourages student action (category 2), or makes constructive interpretations of student feelings or attitudes (category 1). These statements express teacher authority indirectly through their support of selected patterns of student behavior. The teacher assumes a less dominant posture, and the proportion of student talk increases.

"Early and late in the cycle of classroom learning activities" refers to a sequence of events that occurs again and again in every classroom. First, an intellectual difference or problem is created. Second, the major dimensions of the problem are identified, and third, relationships within the problem are isolated. Fourth, work occurs—such as the gathering of information, the application of a formula, or the trial solution of a problem. Fifth, progress is evaluated and tested. Sixth, the new knowledge is applied to additional problems or interpreted in some meaningful way.

There are, of course, cycles within cycles. The solution of a single problem in mathematics is part of a *unit* of study. Both the problem and the unit follow a similar sequence, although a particular problem may involve only five minutes while a unit of study might last two weeks. In social studies, units of study for the entire class and the shorter, individual tasks illustrate these similar relationships. Observation data were obtained separately for the first two days, middle two days, and last two days of a two-week unit of study in order to make comparisons at different stages of the long cycles. Also isolated were short periods of time when new material was introduced, when homework was corrected, or when reports were given. This was done in order to make comparisons within the short cycles.

In order to understand increased or decreased dependence it is necessary, first, to understand what is meant by dependent behavior. A student who is dependent is concerned primarily with pleasing the teacher. As he works on a problem, he is more concerned about whether or not his method is what the teacher wants than he is with whether or not his method will solve the problem. Dependent behavior in a classroom is recognized when students unnecessarily check their work with the teacher to make sure they are on the right track, check their work before going on to the next step, or solicit teacher approval more often than is necessary. Sustained direct influence by a teacher results in increased compliance, and, when this is maintained over an extended period of time, patterns of dependent behavior increase.

The concept of achievement involves measurable differences in content understanding and skill in problem solving as revealed by pre- and postachievement tests. In comparing the achievement of one class with another, adjustments were made for initial ability and knowledge of the subject matter being studied.

When classes in which achievement is high are compared with classes in which it is low, there are certain predictions that can be made as logical derivations from the hypotheses. Hypotheses 2 and 3 both predict high achievement yet require two quite different patterns of teacher influence. Taken together, more variable teacher influence in high achieving classes ought to be found. Early in a learning cycle, the teacher would use an indirect approach; later the teacher might shift to a more direct approach. This shift in influence pattern is called "flexibility of teacher influence." The sustained restriction of student participation by direct teacher influence is associated with low achievement, according to hypothesis 1; and if it is accurate, the prediction can be made that teachers in these classrooms will exhibit limited flexibility of teacher influence.

Considerable research is related to these hypotheses. H. H. Anderson and his colleagues (1939, 1945, 1946) carried out a series of projects with preschool and primary and elementary school children. This imaginative research effort has produced a series of internally consistent and significant findings. First, the dominative and integrative contacts of the teacher set a pattern of behavior that spreads throughout the classroom; the behavior of the teacher, more than that of any other individual, sets the climate of the class. The rule is that, when either type of contact predominates, domination incites further domination and integration stimulates further integration. The teacher's tendency spreads among pupils and persists even when the teacher is no longer in the room. Furthermore, the pattern a teacher develops in one year is likely to persist in his classroom the following year with completely different pupils. Second, when a teacher has a high proportion of integrative contacts, pupils show more spontaneity and initiative, voluntary social contributions, and acts of problem solving. Third, when a teacher has a high proportion of dominative contacts, the pupils are more easily distracted from school work and show more compliance to, as well as rejection of, teacher domination.

In a study by Filson (1957) direct measures of dependent acts by students were associated with different patterns of teacher influence. His study showed that during a task in which the goal is not clear—a condition similar to the early phases of learning cycles—more direct teacher influence elicited more requests for teacher approval and help during work periods even after considerable exposure to the task. The reverse was true when the teacher used an indirect pattern.

Previous research concerning the ways in which teachers modify their behavior to fit different situations and different students has been conducted by several researchers. Wispe (1951) and Smith (1955) have shown that, when college-age students are classified into different psychological types, each type has a different reaction to the same patterns of teacher influence. Gage (1956) found that elementary school pupils' perceptions of the same teacher were different according to classification of the pupil as tending to seek "affective" or "cognitive" responses from the teacher. J. P. Anderson (1960) demonstrated that junior high school boys and girls have significantly ($p < .01$) different ideal expectations about teacher behavior and that students classified as "dependent-prone" tended to expect different teacher behavior than did the "independent-prone" (at the .05 level of significance). The girls and the dependent-prone preferred a pattern that was somewhat indirect.

In the laboratory experiment, Amidon and Flanders (1961) have shown that dependent-prone junior high school students are more sensitive than average students to differences in patterns of teacher influence and that the dependent-prone learned less geometry when exposed to a rigid, direct pattern of influence than they did with an indirect pattern. In this study, the different patterns were created by the same role-playing teacher.

Bales and Strodtbeck (1951) give support to the notion that different types of verbal communication occur naturally in adult group discussions at different stages of a problem-solving cycle.

Other studies involving the classification of verbal statements by using different systems of interaction analysis have been conducted by Bales (1950), Mitzel and Rabinowitz (1953), Withall (1949), and others.

Each of the foregoing researchers has influenced the Minnesota research studies. By 1958, the following generalizations were established concerning patterns of teacher influence.

First, there is a direct relationship between teacher influence that encourages student participation and constructive pupil attitudes toward the teacher, the school work, and the class activities.

Second, different students will have different attitudes in the same classroom situation. In spite of these individual differences, the first generalization holds.

Third, when dealing with the class as a whole, all the evidence suggests a causal relationship—teacher behavior causes pupil attitudes.

Fourth, a number of different factors affect the patterns of influence teachers use. Among these are: (1) the subject matter being taught, (2) the age and maturity of the students, (3) the instructor's preferred style of teaching, and (4) the nature of the learning activities.

Fifth, all teachers employ a combination of statements, some that restrict freedom of participation and others that expand it. Given an extended period of observation, a fairly stable proportion or balance of indirect and direct statements can be identified for each teacher. This ratio, called the "I/D ratio," is positively correlated with the class average on an attitude inventory.

Teacher Influence and Achievement

In order to test the three hypotheses about teacher influence and achievement listed earlier, certain experimental controls were necessary. First, all teachers in each content area had access to the same instructional materials. They were provided in sufficient variety so that each teacher could use his natural style of teaching. Second, the procedure for testing achievement gain permitted the testers to adjust for initial ability and knowledge. Third, the teachers in each subject-matter area were selected from two large urban areas so that they would be representative of the wide range of teacher influence patterns found in the parent population. Fourth, a schedule was carefully worked out so that members of the research staff could administer the tests and make the necessary observations at the beginning, middle, and end of the two-week units of study. Fifth, mathematics and combined English-social studies were chosen in order to test the hypotheses under very different conditions—mathematics with its tight, logical, inductive-deductive reasoning, and social studies with less exact reasoning systems.

A choice was soon evident in designating the dependent and independent variable. High and low achieving classes could be isolated and then compared with the teachers' influence patterns, or teachers could be grouped according to their patterns of influence and then their achievement compared. If the hypotheses were valid, of course, there would be no difference in the results. However, designating teacher influence as the independent variable and then predicting achievement was more consistent with the way that the hypotheses were stated.

CONTROL OF THE INDEPENDENT VARIABLE The job here was to identify the pattern of influence used by each teacher, to find out which teachers varied their proportion of direct and indirect influence within cycles of learning, and to group teachers according to the characteristics of their influence patterns. The research tool used was interaction analysis.

The first problem was to establish the reliability of the observers and maintain this reliability during the three-month data-collecting period. All observers began with twelve to eighteen hours of training before the first classroom visit. Preliminary training made use of tape recordings for which exact distributions were already known. Later, the observing team made use of one-way observation booths that permitted observers to talk out loud to one another as they classified statements in the adjoining "live" classroom. These procedures provided sufficient reliability, given intelligent graduate students who had previous counseling or teaching experience.

Once established, the maintenance of reliability required weekly discussions about unusual categorizing problems and spot-reliability checks with a fellow observer. Reliability may, of course, deteriorate unless maintained by additional training. A technical discussion of these procedures can be found in *Interaction Analysis: A Manual for Observers* (Flanders, 1960b).

An analysis of the statements of the sixteen mathematics teachers and fifteen social studies teachers (one was dropped due to unusual circumstances) suggested that three groups of teachers would provide the most useful comparisons. In each subject, the teachers were classified as most indirect, average, and most direct. This division was based on a restricted proportion of indirect to direct statements, known as the i/d ratio, computed for each teacher for all hours of observation.[3] The composite distributions for the mathematics teachers are shown in Figure 2, social

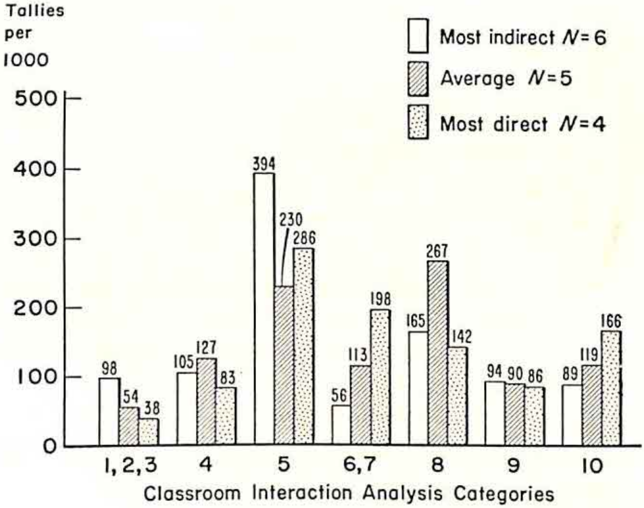


Fig. 2. Interaction-analysis data from most indirect, average, and most direct social studies teachers.

---

[3] The i/d ratio makes use of all seven teacher categories. Referring to the categories listed above, i/d = (1 + 2 + 3 +4) divided by (5 + 6 + 7). The i/d ratio ignored categories 4 and 5, asking questions and lecturing; so that i/d = (1 + 2 + 3) divided (6 + 7). The i/d ratio is preferred in this application because it is independent of communication patterns (like drill) that are unique to subject matter.

studies in Figure 3. In comparing these two histograms with Figure 1, it can be seen that the differences between high and low scoring classes in New Zealand were similar to the group differences in social studies and mathematics even though there were some basic differences in the over-all patterns.

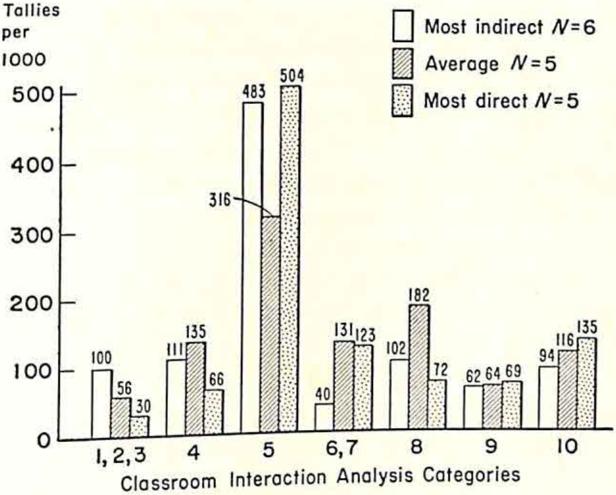


Fig. 3. Interaction-analysis data from most indirect, average, and most direct mathematics teachers.

Most seventh-grade social studies classrooms are quite different from eighth-grade mathematics classrooms, yet the similarities between Figures 2 and 3 are quite remarkable. By comparing the bars above categories 1, 2, and 3, with 6 and 7, it is obvious that the i/d ratios of the three groups form a rank order that is consistent with the method of grouping the teachers. It was something of a surprise to find that the composite i/d ratio of the most indirect mathematics teachers was larger than the similar ratio for social studies teachers.

There were a number of characteristics that distinguished the most indirect teachers from the most direct in both social studies and mathematics. First, indirect teachers were more alert to, concerned with, and made greater use of statements made by students. These teachers went beyond mere clarification and acknowledgement of student ideas; they skillfully integrated student ideas into the content discourse of classroom communication. The teacher's clarification and development of student ideas over periods longer than three seconds[4] occurred more frequently by a factor of eighteen when the most indirect social studies teachers were compared with the most direct. For mathematics, the same factor is five.

Second, the most indirect teachers asked longer, more extended questions, and did this about four times more frequently than did the most direct teachers. Short questions usually elicited short student responses; long questions, long student responses. Ideas were dealt with in greater detail in the most indirect classrooms.

Third, the most direct teachers had more discipline problems and found it necessary to interrupt giving directions in order to criticize students three times more often than did the most indirect. The most direct teachers gave longer and more involved directions and often had to repeat directions. Students more often tended to question or even to resist the directions given by the most direct teachers.

Some readers may conclude, at this point, that the most direct teachers in this study had the misfortune to be dealing with difficult classes at the time of the project. While it is true that some classes are more difficult than others, this did not affect research results to any great degree. In the first place, some of the most indirect teachers were assigned troublesome students as a regular policy. It was the testers' impression that sustained and rigidly employed patterns of above-average direct influence invited student resistance. This observation is supported by Anderson, Brewer, and Reed (1946). They found that teachers soon established, in the second year, the same or similar patterns observed during the first year in spite of different pupils. Anderson was observing primary classes, however; and there can be no doubt that, at older age levels, classroom interaction is the product of both teaching style and pupil reaction. The point, here, is that the teacher has the power, the experience, and the professional responsibility to establish the quality of the interaction.

---

[4] Interpretations from interaction analysis data made in this section are based on "matrix interpretation." The matrix has not been described here due to lack of space. An extended treatment of this method of tabulating and interpreting interaction data can be found in *A Manual for Observers*, Flanders (1960b).

The i/d ratios used to divide teachers into the most indirect, average, and most direct groups, as well as the profiles shown in Figures 2 and 3, were based on data combined for all hours of observation. In the mathematics classes, there were six one-hour periods, and in combined English-social studies classes, six two-hour periods. However, during each classroom visit, a number of different learning activities occurred that could be clearly identified by the observer. In mathematics, these time-use categories included: (1) *routine* administration—such as taking roll, distributing or collecting materials, and so forth; (2) *evaluation* of the products of learning—such as discussion of homework, test results, or student reports; (3) introducing *new material*—such as a class discussion of a new formula (or new facts about a country); (4) *general discussion* of old material or current problems; and (5) *supervision of seat work*—periods when students were busy doing homework or projects. The same five time-use categories were used in social studies plus a sixth not involved with mathematics, namely (6) *teacher-pupil planning*—or class discussions about the organization of work.

Inasmuch as the interaction analysis data were tabulated separately for each time-use category, providing it appeared, it was possible to calculate separate i/d ratios for each type of activity. Later will be shown the extent to which changes in i/d ratio across different learning activities supported the hypotheses of this study. However, a preliminary question to be asked, now, concerns the degree of variability in range of different i/d ratios when the most indirect is compared with the most direct. In Figures 4 and 5, the frequency of different i/d ratios is shown for social studies and mathematics. Notice that the class intervals for each bar in the histogram are not equal.

All of the i/d ratios of the most direct teachers fall within the range of 0.01 to 2.0, with the majority below 0.4. For the most indirect teachers, the range for social studies is 0.01. to 18.0. For mathematics the range is 0.01 to 11.0. The majority of ratios for the most indirect fall above one. The distribution of i/d ratios for the two average groups fall midway between the two extreme groups and are not shown on the histograms.

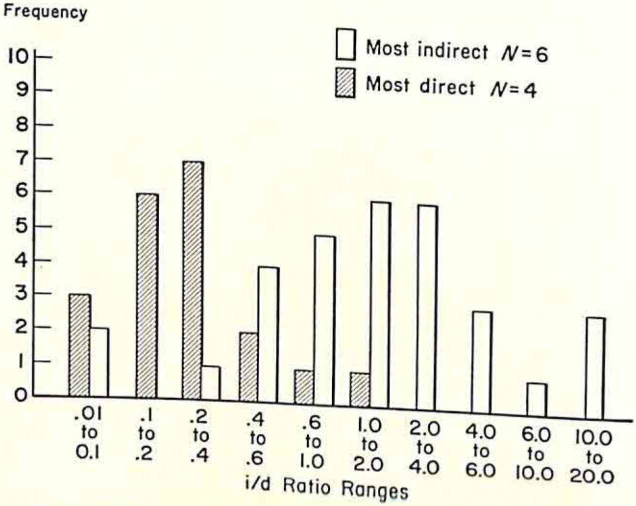


Fig. 4. Distribution of i/d ratios for different classroom activities in social studies.

These data neither support nor reject the three hypotheses of the study, but they do indicate that we can expect greater flexibility of teacher influence among the most indirect and less flexibility among the most direct. Clearly, the most indirect are making more dramatic shifts in their patterns of teacher influence as the classroom learning activities change from one time-use category to another.

RESULTS IN TERMS OF ACHIEVEMENTS The first step in the analysis of achievement data was the making of adjustments for initial ability and differences in IQ. Adjustments were made by analysis of covariance based on pretest scores only. Since IQ was positively correlated with both pretest and posttest scores of achievement, the extra work of including IQ in the regression equation was not justified.

Next, in placing all teachers in each subject area in a rank order list according to over-all i/d ratio, it was noticed that a large break

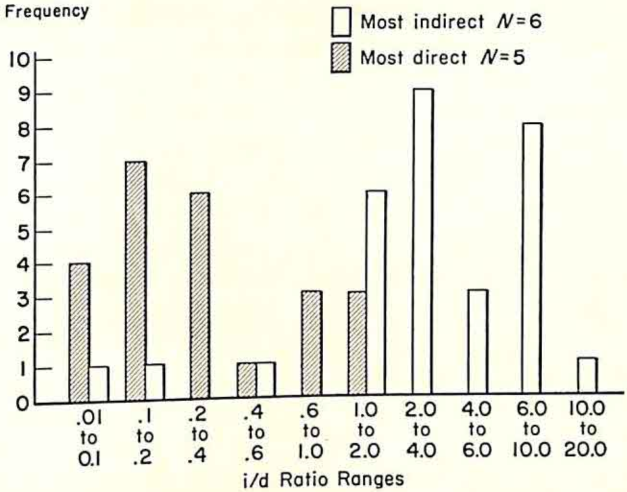


Fig. 5. Distribution of i/d ratios for different classroom activities in mathematics.

occurred between the seventh and eighth-grade teachers in mathematics and that the social studies rank order progressed evenly. Consequently, a comparison of adjusted achievement scores was made for seven indirect versus nine direct mathematics teachers; a similar comparison was made for eight indirect and eight direct social studies teachers until one of the teachers was dropped, as indicated earlier, making a seven versus eight split. The results of these comparisons—not shown here—indicated superior achievement for the indirect classes, when compared with the direct, at a level of significance beyond 0.01.

The results of the analysis of adjusted achievement scores for the most indirect and most direct teachers are shown in Table 1. Achievement was significantly higher in the most indirect classes in both social studies and mathematics.

RESULTS IN TERMS OF DEPENDENCE AND OTHER ATTITUDES The testers were unable to develop a satisfactory system for observing spontaneous acts of dependence in the field. However, a long form of student attitude inventory was administered similar to the attitude tests used in the research prior to 1958. This long form contains items that indicate a student's perceptions of the teacher, the class, and the learning activities. Many items in the inventory were intended to be direct measures of independent perceptions. Additional items were keyed in a direction logically related to independence, although they dealt with the teacher, motivation, or other factors in the classroom. Thus, a high class average indicates favorable and constructive attitudes, in general, with minimum dependence on the teacher.

TABLE 1

Adjusted Achievement Scores for the Most Indirect and Most Direct Social Studies and Mathematic Classes

	Mathematics				*Social studies*			
			N				*N*	
Teaching Style	Adjusted Mean	Variance	Students	Teachers	Adjusted Mean	Variance	Students	Teachers
Most Indirect	28.9	42.3	156	6	36.4	20.4	156	6
Most Direct	24.9	33.3	90	5	31.0	31.8	92	4
Critical Ratio	5.00		$p < .01$		7.73		$p < .01$	

As with the achievement data, an analysis of attitude scores of all classes was prepared to reveal high scores ($p < .01$) for the indirect patterns of teaching. These differences were even more significant when the most indirect and most direct classes in social studies were analyzed separately. In mathematics, however, the degrees of difference between the most indirect and the most direct teachers were only slightly greater than for all teachers. The data on which these statements are based can be found on page 91 of the U.S. Office of Education Terminal Contract Report (Flanders, 1960c).

In this field experiment, just as in the earlier studies not involving achievement, the most constructive and independent attitudes were found to be associated with the most indirect patterns of teacher influence. The relationship between these attitude measures and measures of adjusted achievement is expressed by a correlation of r-.44 for mathematics and r-.45 for social studies.

RESULTS IN TERMS OF THE MAJOR HYPOTHESES The theoretical considerations underlying the three hypotheses stated earlier have been developed in greater detail elsewhere (Flanders, 1960a). The characteristics of the most indirect and most direct teaching patterns already have been described. It is now known that, in the indirect classrooms, students learned more and possessed more constructive and independent attitudes than in the direct. Furthermore, the indirect teachers were more flexible than the direct teachers in that they made more dramatic changes in their patterns of influence in the various time-use activity categories.

It is difficult to describe these results without leaving the impression that there *is* such a thing as an indirect teacher and a direct teacher. These labels were used to refer to the groups of teachers compared in these various research projects. It cannot be overemphasized, however, that there is no direct or indirect teacher. All teachers are either indirect or direct over only very short periods of time. Every teacher, over long periods of time, blends direct and indirect acts into some kind of balance. When all of the acts observed for a group of teachers are combined into a composite pattern, a comparison of teachers tending toward either direct or indirect teaching patterns reveals differences that total only 20 to 30 percent in certain crucial categories. These are shown in Figure 1.

Even though the general differences between indirect and direct teaching are interesting and of practical importance to the profession, further analysis is required to support the three major hypotheses. These hypotheses predict certain consequences for pupil dependence and achievement from particular patterns of teacher influence. The rationale supporting these hypotheses involves the following logical steps.

First, the hypotheses describe shifts in teacher influence during the early and late phases of learning cycles. Therefore, these hypotheses could be tested only if some kind of shift took place over a period of time. Inflexible teaching patterns would not provide evidence one way or another.

Second, by isolating the most indirect teachers—who were shown to be more flexible—it would be possible to study the shifts these teachers made to see if they were consistent with the hypotheses. It was expected that the most direct teaching patterns—in which shifts of influence had been shown to be very small—would not support these hypothesized trends significantly (perhaps they would even make small shifts in the opposite direction).

Third, if the previous two steps proved fruitful, it was then expected that learning would be significantly higher in those classrooms where theoretical predictions were supported; significantly less learning was expected in those classrooms where the hypotheses were not supported.

A comparison of the first two days, middle two days, and last two days of the two-week units of study provides evidence of changes in teacher influence patterns over the long-range cycles (see Figure 6).

Two histograms for social studies and two for mathematics are shown in Figure 6. The percentages of total class time spent in either "expanding" or "restricting" activities are shown at the left. All situations observed that obviously permitted student freedom of action were classified as expanding. Those that obviously restricted student freedom of action were called restricting. Unless the observer's notes were clear and unequivocal, the interaction record was assigned to a neutral category for the time period coded. Situations that, typically, were classified as expanding included teacher-pupil planning with high student participation, discussion designed to clarify the students' point of view, periods when the students directed their own discussion, and so on. Situations that, typically, were classified as restricting included assignment giving by the teacher, the telling of students how to do their work, and so on.

First, look at the absolute differences in height of the clear bars and the shaded bars in all four histograms of Figure 6. The social studies and mathematics expanding histograms show that indirect teachers provided significantly more expanding activities at all stages of learning except for the last two days of mathematics units. In this one instance, the trend is consistent but not significant. These absolute differences are smaller for restricting

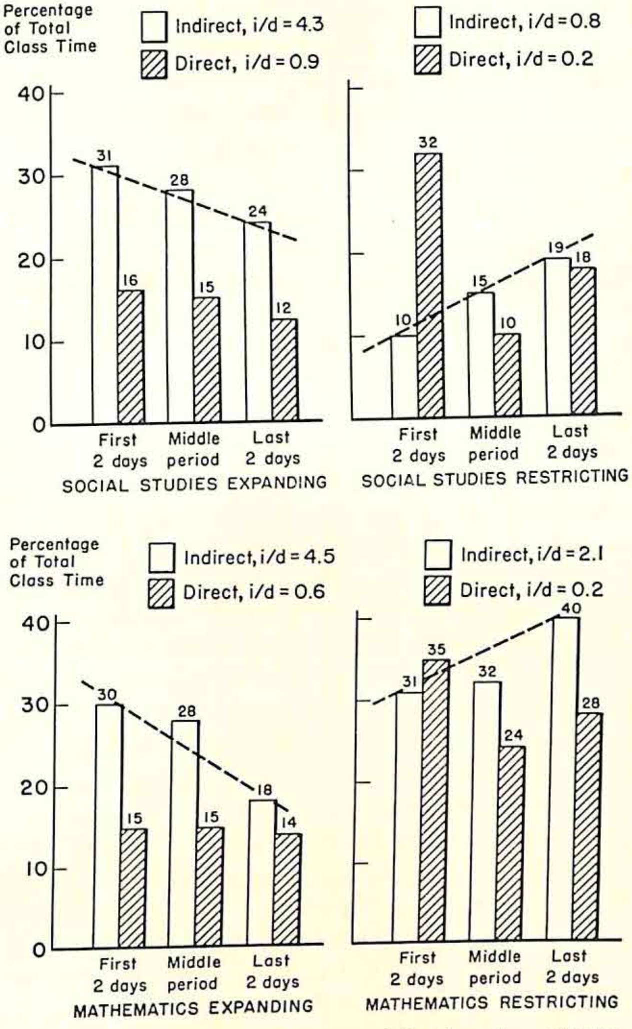


Fig. 6. Shifts in expanding and restricting learning activities.

activities as shown in the histograms. Here, all differences are statistically significant except for the first two days of mathematics and last two days of social studies.

The fact that expanding activities were significantly higher for the indirect teachers and restricting activities were not always consistently higher for the direct teachers confirmed results in the earlier studies. On several occasions, the differences between direct and indirect teachers did not appear in the direct aspects of the teaching-learning process but were present in the indirect aspects. This is interpreted to mean that indirect teachers can be direct, but direct teachers cannot be equally indirect.

Second, the trends predicted by the hypotheses are shown by the dash lines in each histogram. The decreasing use of expanding activities is clearly shown for the indirect teachers and is nonexistent, for all practical purposes, for the direct teachers—as shown in the histograms of expanding activities. The tendency to increase restricting activities is not only present for the indirect teachers (shown in the histograms), but the opposite trend appears to approximate better the shifts of the direct teachers.

Not only did the use of time support the hypothesized trends, but the i/d ratios, shown above each histogram, added further support. For example, expanding activities in social studies occurred more frequently and with a much more indirect pattern of teacher influence in the indirect classrooms—the direct classroom providing less time for this kind of activity, and, when it did occur, the teacher was much more direct.

Altogether, the four histograms support the three hypotheses. Hypotheses 1 and 2 predicted the increasing use of direct influence for increased learning and decreased dependence. This trend was followed by the indirect teachers; the opposite trend was followed by the direct teachers. Hypothesis 3 predicted indirect influence early in the cycle for increased learning and decreased dependence; these were clearly present in the indirect classrooms.

Finally, turn to cycles that occur during a single observer visit. A teacher conceivably can introduce some new material, discuss it with the class, and assign it as work—all within a period of thirty minutes or less. The dramatic changes in influence by the indirect teacher for different situations, and the smaller adjustments by the direct teachers, are shown in Table 2. The i/d ratios for the six different time-use categories in social studies and the five in mathematics show that indirect teachers made greater changes.

TABLE 2

Median i/d Ratios for Direct and Indirect Teachers

Subject	*Type*	*Planning*	*New material*	*Discussion*	*Work*	*Evaluation*	*Routine*
Social Studies	Indirect	5.14	3.25	1.72	0.84	0.93	0.30
	Direct	1.31	0.23	0.51	0.03	0.20	0.17
Mathematics	Indirect	—	3.64	6.83	2.24	2.24	0.59
	Direct	—	0.72	0.21	0.15	0.25	0.14

SUMMARY AND CONCLUSIONS The three major hypotheses were supported by an analysis of data from interaction analysis, attitude, and achievement scores in seventh grade social studies and eighth grade mathematics. Teachers who were able to provide flexible patterns of influence, by shifting from indirect to direct with the passage of time, created situations in which students learned more. The students of teachers who were unable to do this learned less.

There seemed to be four essential elements of teacher influence in the classrooms in which achievement and attitudes were superior. First, the teacher was capable of providing a range of roles, spontaneously, that varied from fairly active, dominative supervision, on the one hand, to reflective, discriminating support, on the other hand. The teacher was able not only to achieve compliance but to support and encourage student initiative. Second, the teacher was able to control his own spontaneous behavior so that he could assume one role or another at will. Third, he had sufficient understanding of principles of teacher influence to make possible a logical bridge between his diagnosis of the present situation and the various actions he could take. Fourth, he was a sensitive, objective observer who could make valid diagnoses of current conditions. All of these skills, which seemed to characterize the most successful teachers, were superimposed upon a firm grasp of the subject matter being taught.

Teachers in the studies who failed to approach these requirements appeared to be restricted to a limited number of roles producing a rigid sequence of actions. There was little variation from one classroom situation to the next. In particular, the inability to expand or restrict the freedom of action of the students through one's self-control of verbal influence seemed to characterize the unsuccessful teachers.

All of these conclusions are made from composite bits of teacher behavior for two groups of teachers. They are not conclusions based on the actions of a single teacher in an expanded case study. These generalizations also are restricted by what is best called "the limitations of current practice." Were it possible to give one year's training to a group of currently superior teachers, accentuating their awareness of influence patterns and helping them to develop even more self-control, the performance differences might be increased to become even more striking than those found in the present study.

Some Observations Concerning the Evaluation of Teacher Competence

The research activities and conclusions presented in preceding sections of the chapter involved past efforts. They summarize what happened in a series of projects concerned with the study of teacher influence; now, to a contemporary interpretation of these past events in terms of the purpose of this book.

The Problems of Evaluating Teaching

There are those who say that teaching is an art; and certainly no one can deny that an exceptionally fine lesson has much in common with an artistic performance. Most artists would be the first to point out, however, that arduous, lengthy practice and attention to technical skill were prerequisite to a particularly fine performance. This last observation seems inconsistent with the old expression that teachers or artists are born, not made. The education and training of a teacher involves a science to the extent that there are logical relationships among what a teacher does, his own understanding of what he does, and his ability to organize these relationships into orderly principles.

Today, we are closer to a scientific understanding of teaching than ever before, and we are approaching a theory of teaching. Much more is known about teaching than a close study of current teacher preparation programs might suggest. The day is close at hand, if not already here, when prospective teachers can no longer be exposed to some facts about individual differences, some theories of learning, child growth and development, and so forth, and let the entire burden of translating this knowledge into teaching action fall on the practice-teaching experience.

In current situations, the evaluation of teaching is concerned not only with the improvement of current practices but with overcoming deficiencies in teacher preparation. In this section, some of the difficulties existing in current conditions will be dealt with briefly.

ATTAINING THE COOPERATION OF TEACHERS One problem met in both the evaluation of teaching and research on teacher influence involves the requirement of cooperation from the teachers. In this section, the procedures used for soliciting teacher cooperation will be reviewed, as they may be equally useful when teaching is evaluated.

Teachers will cooperate if they agree with the purposes of evaluation. They will resist if they mistrust these purposes. The decision to cooperate rests on *their* perceptions, not those of the administrators or researchers. The rule followed—when talking with the top administrators, the building principals, and the classroom teachers—was to approach each group as if the others did not exist. The principals knew that the project had cleared the superintendent's office, and the teachers knew that the principals had been consulted. However, it was emphasized that these clearances were for the purpose of *approaching* the next group, and that each group made its own decision to cooperate.

Stressed in meetings with administrators was the importance of creating an in-service training committee and a research steering committee (one for each project). For the in-service training project, the majority of committee members were classroom teachers; for the field experiments, there were two parents, two classroom teachers, two building principals, and two representatives of central administration. Each committee's functions were: (1) to review every paper-and-pencil test, observation technique, and experimental procedure; (2) to discuss these matters until they were understood, and, then, either to endorse them or to incorporate appropriate modifications so that they could be endorsed with the committee decision being final; (3) to act as consultants to the research team—this involving research plans, implementation of plans, and possible approaches to the teachers; and (4) to act as a contact point for any teacher, pupil, parent, or administrator who became upset by the research activities.

The question each member of the committee asked himself, after reviewing research plans, was whether or not he could honestly say that he approved. It was important for the parents on the research committee to agree that their children be experimental subjects; it was necessary that the teachers be willing to participate, or to have their classes participate if asked; that the principals be willing to have the project in their schools; and it was important that the central office representatives approve from a system-wide point of view.

The steering committees were formed *before* talks with building principals and classroom teachers—in fact, the committees helped to plan these talks. In order to develop the field experiments, the author and associates met first with teachers in small groups of fifteen to twenty-five at convenient, nearby schools. At most of these meetings, about one half of the teachers were willing to participate after listening to the explanation. A higher percentage required additional discussion. The lowest percentage of volunteers occurred at one meeting during which they questioned closely about whether or not this research was likely to be a first step toward merit pay evaluations.

The in-service training project was first described at two school-wide meetings. Next, all teachers were provided with a written brochure giving more detailed information. The day before teachers were to volunteer, a member of the research staff was available all day in the teachers' common room for private question-and-answer sessions. Many questions seemed to be directed only toward becoming better acquainted; but considerable concern was expressed about the purposes of the project, safeguards to prevent data from reaching the hands of administrators, and the amount of time the project would take. Even these informal contacts were not enough, however, to establish trust and cooperation among all interested teachers.

One incident sensitized the research team to the need for additional informal and private channels of communication. A teacher called a member of the in-service steering committee to indicate that the volunteer aspect of the participation did not exist. Those who did not volunteer automatically would be identified, and the administration would then classify them as not interested in professional development. This was discussed until the following solution was suggested. The teacher could volunteer for the in-service research project to show "professional interest," but when the hours for meetings were announced, he then could ask to be excused because of "a conflict of time caused by previous commitments." By means of expressions of disappointment, he could avoid being classified as not interested in professional growth. This alternative was sent back along the same informal channel of communication, and, when other teachers heard about it, acceptance of the project probably was increased.

The following conclusions were reached: (1) the study of teacher behavior is acceptable to about one half of the teachers—the rest require special approaches; (2) physical access can be provided by administrators, but genuine psychological access can come only from the teachers (for example, full access is not present when a teacher pretends something for an observer's benefit); (3) teacher control is provided by approval and veto powers on a "steering committee"; (4) not all teachers believe explanations of purpose—they require personal and private conversations during which they can speak freely; (5) a teacher is in the best position to solicit another teacher's cooperation; (6) the reduction of fear, suspicion, and skepticism occurs with frank and forthright communication among all concerned.

These observations probably apply, to some extent, to any program aimed at the evaluation of teaching.

EVALUATION OF TEACHING IN ORDER TO IMPROVE INSTRUCTION

In the final analysis, all evaluation of teaching has the ultimate purpose of improving instruction. A merit pay system seeks to do this by a special, formalized system of monetary and prestige rewards. In the in-service training program, rewards were individually created and recognized. Most of the teachers learned more about themselves, about their classroom behavior, and about their students' reactions. This was rewarding in itself. More than one half of the participating teachers were able to use this information to make changes in their classroom behavior; this, too, was rewarding. This project is believed to be the first attempt in education to judge the effectiveness of in-service training through a systematic analysis of spontaneous teacher behavior in the classroom. Heretofore, shifts in attitude, knowledge of teaching methods, or intentions—all measured by paper-and-pencil instruments—were used as evidence of progress before and after training. Just how much attitude change was translated into action was usually never known.

The approach began with the observation that most teachers have already established values in educational philosophy. These are not necessarily related to their behavior. Attempting to change these values is not the most direct approach to changing behavior. Qualified teachers have passed examinations that test their judgment of what "ought to be done in the name of good teaching"; they are very good at identifying democratic values. They have developed diagnostic skills at the paper-and-pencil test level and perform effectively at this level. However, the real issue is the relationship between the teacher's intent and his action. This issue was stressed in these studies by asking two questions: "What do you want to accomplish?" and "What did you do in order to achieve these ends?"

The most effective changes in methods of instruction occur when a teacher can compare what he wanted to accomplish with a nonthreatening, objective summarization of his spontaneous behavior. Using proper procedures, the teacher can make his own discoveries and reach his own conclusions about what changes would reduce any discrepancy between intent and action.

These observations have suggested a few principles found helpful in the design of in-service training. First, do not consult with a teacher after a single visit because the teacher will initiate a series of defensive questions designed to obtain value judgments from the observer (for example, "What did *you* think of the lesson?"). Always schedule consultations after two or more visits so that comparisons can be made between different class sessions (for example, "Why was there more student participation in the first-period class?").

Second, the evidence and the criteria to be applied to evidence must be available equally to consultant and teacher. This means that all data-gathering procedures must be understood by the teacher. In the case of interaction analysis, teachers should have some experience as observers before being observed themselves. It means also that the data to be used in a particular consultation should be available to the teacher before the interview with a consultant. The teacher will need time to study the material in order to organize his own reactions.

The data to be collected for a consultation should be determined in advance by questions asked by the teacher. Questions to determine if something was good or bad and if certain changes ought to take place are poor questions at early stages. Questions should focus on what happened and how one situation differs from another.

Third, the major purpose of evaluating instruction is to create a sense of inquiry and experimentation in which one variable is a change in the teacher's behavior. A simple sequence is to: (1) speculate about what might happen if the teacher acted first one way and then another in two somewhat similar situations; (2) try the variations with adequate data collection; and (3) analyze the results. With the right atmosphere, pairs or trios of teachers soon form teams and become more independent of consultant help.

The development of a sense of inquiry that includes experimentation with one's own behavior is the most important product of in-service training activities. Yet this emerges slowly as a characteristic of a school program. There are many restraints and inhibitions within a faculty to block the development of a sense of inquiry. Often, the teachers who could benefit most find it very difficult to experiment with their own behavior.

Fourth, creative tools for gathering a wide range of data is necessary to accommodate the range of questions teachers are prepared to ask. Teachers often ask two kinds of questions about pupil reactions and attitudes. The first has to do with long-range attitudes, such as those measured by attitude inventories. Second, they often like to collect opinions about a single lesson, a unit of study, or even a particular incident. Questions about teacher behavior were encouraged in our in-service training project, and they required complex systems of observation, such as interaction analysis.

There is no limit to the ways that data can be collected. One art teacher rated art products according to the degree to which the pupil copied an example executed by the teacher. This was done as a measure of pupil dependence on teacher influence. A mathematics teacher designed a "standardized" problem-solving experience for small groups of pupils as a measure of transfer from textbook problems to unique and unusual problems. Each teacher operates at his own level of sophistication in measurement, and this level changes with experience and training.

EVALUATING TEACHING IN ORDER TO CLASSIFY TEACHERS Any system of merit pay, or other means of giving special recognition to individual teachers, must solve the problems associated with "single case decisions." These problems can be illustrated as follows.

A counselor collects test information about a single student to produce a profile of aptitudes and interests. Studies indicate that 80 out of 100 students with scores similar to those of the student succeed in the field of medicine. The crucial issue in making a recommendation involves determining if this particular student will be among the twenty who will fail or the eighty who will succeed.

The second illustration involves the assignment of grades by a classroom teacher. Suppose a teacher were blessed with achievement tests that were perfectly reliable and valid. Furthermore, the tests adequately sampled the major content objectives of learning. The test scores then would be valid indexes of each student's achievement—a condition that is approached only in practical situations. The fact is that even though a teacher can "scale" individual achievement accurately, he still faces an arbitrary, politico-social decision as soon as he assigns a grade. By what system of reasoning is one level of performance a failure, another set aside as an "A"?

These two problems are present in any system of classifying teacher performance. There will always be the question of whether or not the decision taken in a single case is the right one. Similarly, there will always be philosophical issues present in the use of evaluation data to classify teachers into categories of ability—and these will have personal consequences for the teacher and the school system.

Each person reading this material will have his own opinions about these problems, but there are some fairly objective observations that should be made.

First, in the studies of teacher influence, it was found that those teachers whose students had less constructive attitudes and who learned less than others were much more alike, as a group, than were the teachers classified as superior. It is easier to identify poor teaching than it is to identify superior teaching. The characteristics of the former are more consistent than the latter.

Second, the research on teacher influence compared groups of teachers, and the conclusions reached were based on comparisons of two hypothetical composite teachers, whose acts of influence were, in fact, all the acts of several similar teachers. None of the comparisons were based on a single case. In addition, data were collected in situations that were carefully constructed in order to make them as similar as possible. It is interesting to note here that some of the correlations between interaction-analysis data and scores of pupil attitudes collected in the heterogeneous classrooms of the in-service training project were lower than those collected in field experiments. That is, the interaction-analysis data in a girls' physical education class is drastically different from that in a science class, which is, in turn, different from general shop.

One additional incident may be mentioned. While studying teacher influence patterns in New Zealand, it was possible to identify a rank order of ten teachers, based first on the attitude scores of the pupils, and second, on interaction-analysis data that distinguished the top five from the bottom five in a fashion consistent with theories about teacher influence. Just before the author left New Zealand, the ratings that school inspectors had given to these teachers over the past few years became available. School inspectors in New Zealand give teachers a rating along a single scale of teaching effectiveness. This affects promotion and determines who receives desirable positions when they are open. The relationship between these ratings and the research data was quite low. Two out of five of the top teachers had received below-average ratings. Three out of five of the bottom teachers had received average or above-average ratings. An inspector normally visits a teacher for a half day and then writes a report based on a check list of classroom characteristics. The two systems produced contradictory results.

Third, the last issue in classification to be mentioned concerns the consequences of making an error. Before a surgeon will agree to a serious operation, he insists on evidence that is sufficiently conclusive to justify the risk. Standards nevertheless vary among surgeons. The same can be said about courts of law that must judge a man insane or guilty of some crime. Educators who choose to classify teachers according to their teaching ability must resolve issues inherent in the following questions: How large is the error of classification? What are the consequences of making a mistake? Will the entire process achieve the original goals of the evaluation program?

Finally, one finding from the in-service training project runs contrary to the thinking of some researchers who study the traits of teachers as an approach to the evaluation of teaching. Storlie

(1961) found that changes in classroom behavior resulting from in-service training, as measured by interaction analysis, were not correlated with eight different personality measures. These measures were selected as the most promising from Cattell's sixteen-factor personality tests, the eight scales in the Runner Questionnaire, and the Minnesota Teacher Attitude Inventory, a total of twenty-five scales. He concluded that the measures of personality he investigated did not help to predict which teachers would benefit most from in-service training. Another study (Davies, 1961) indicated that relationships between personality measures and pattern of influence, before in-service training, are sufficiently obscure so that an analysis of variance produced nonsignificant results. Perhaps a more sensitive analysis, such as the Johnson-Neiman technique used by Bowers and Soar (1961), may be more useful. Davies concluded, however, that so many factors influence the teacher's behavior that measures of teacher personality account for very little of the variance.

A Point of View

We live in a society in which genuine, helpful evaluation is quite uncommon. Notice how uncomfortable we feel when someone pays us a compliment and how defensively we respond to almost any kind of criticism. As a group, teachers are isolated from systematic information about their own behavior. However, research indicates that teaching performance depends upon the range of behavior a teacher can produce, the self-control required to provide particular patterns of influence, a teacher's sensitivity in diagnosing the requirements of the moment, and his ability to predict the consequences of alternative actions.

The scientific study of teaching is so immature that, at this time, a particular pattern of teaching cannot be advocated as the most successful. Considerable improvement probably can be brought about by any program that provides reasonably intelligent teachers of average emotional adjustment with systematic information about their classroom behavior. If the information at hand increases, teachers will make modifications that can be tested, and, if satisfactory, adopted as improvements.

THE TEACHERS' BILL OF RIGHTS First, each teacher needs the freedom to develop his own unique style of teaching. While teachers of the same subject matter and age level will often have much in common, there will always be unique differences as products of teacher-pupil interaction. No one can squeeze all teachers into the same mold.

Second, only the teacher himself can change his methods of teaching. Teachers can be told how to teach; they even can be ordered to teach in a particular manner; but stable changes in patterns of teacher influence—changes that result in improvements—are self-motivated. Surveillance is somewhat impractical. Only the changes that make sense to the teacher are maintained. A supervisor who helps a teacher to discover some better method for himself is of real help.

Third, helping a teacher to change his methods of teaching involves emotional as well as intellectual problems. Awareness of new material, access to new instructional resources, development of a new idea are all important in making changes. However, readiness for change must be considered as well as the fact that trying out new methods arouses emotional reactions in most teachers. Reactions that lead to constructive and positive viewpoints are essential for stable innovations. Reactions that are less than positive must be recognized, analyzed, and tested in situations free of threat if a realistic approach to innovation is to be taken.

Fourth, efforts toward change require social support, as well as intellectual stimulation, from one's peers. Two or three like-minded teachers, who have developed a sense of experimentation in their approach to teaching, may form a highly stable unit of change.

Taken all together, these four points suggest that any program for the improvement of teaching must be a voluntary, teacher-directed enterprise. Coercion and administrative fiats are not likely to succeed. Pressures for change can be generated in voluntary peer association, but compliance alone is not likely to lead to stable innovation.

THE NONTEACHERS' BILL OF RIGHTS Nonteachers include supervisors, administrators, members of school boards, and the general public.

First, the long-established right of privacy within the four walls of the classroom cannot be used indiscriminately as a shield of academic freedom to block all possible approaches to change. Improvement of instruction takes place in the classroom, if it is to occur at all; and evidence of change must be available to all members of an improvement team, teacher and nonteacher alike.

Second, the fact that teaching is a complex social process, hard to define and evaluate, does not mean that all evidence is useless simply because it is incomplete. The tools and techniques for establishing criteria of teaching effectiveness are crude, but they can be improved only by further experimentation and development.

Third, nonteachers have a right to expect the teaching profession to take an active part in the development of successful methods for evaluating teacher effectiveness. The mark of a true professional is an active concern for developing criteria for professional effectiveness.

S.C.E.R.T., West Bengal
Library
Calcutta
25/3, B. C. Road

8

HOWARD A. ROSENCRANZ *and* BRUCE J. BIDDLE

The Role Approach to Teacher Competence

It is neither profound nor astute to observe that the field of teacher competence, so long a focal point of interest for so many investigators, has produced few results acceptable either to teachers or to those who judge them. Examination of this paradox is essential in defining the problem at hand.

Why this disparity between effort and results? A number of explanations have been suggested by commentators:

1. No past research has been adequate in methodology.
2. So many studies exist that usable and appropriate methods are lost among the welter of poor research results.
3. Teacher behaviors are so complex that the behaviorial sciences, in their present stage of development, cannot deal with the problem.
4. Research on such a practical problem as teacher competence is bound to fail because only "basic research" produces real breakthroughs.
5. Teachers are constitutionally and professionally opposed to having their performances evaluated and refuse to cooperate

with research in this field and do not allow a measurement of competence by established means.

As impressive as these suggestions may be, it seems to the authors that none can be accepted in any direct sense. While there is much poor research in the teacher competence area, many well-designed empirical studies are reported in the literature. Indeed, summaries of competence research such as that by Morsh and Wilder (1954) cannot but impress one with the variety of measurement devices, the objectivity of approaches, the varieties of samples, and the sophistication of statistical techniques. Competence research may be segmental, piecemeal, and noncumulative, but it has certainly been adequate.

In turning to the "adolescence" of the social sciences, it must be granted that teacher behavior is a very complex subject and that less is known about human behavior, generally, than is known about natural science phenomena. But many "practical" problems in human behavior are under effective attack today, among them problems of mental health, reduction of prejudice, and motivation research—to name but a few. It is unreasonable to assume that teacher behavior is of a different order of complexity than these other—more successful—fields. It also should be pointed out that some of the most rapid advances in these fields come as the result of "practical" research projects, such as the one that produced *The Authoritarian Personality* (Adorno, *et al.*, 1950). It should be assumed that both basic and applied research are needed and that each offers something the other cannot provide.

Finally, factors are undoubtedly working against the acceptance of a method of competence evaluation by educators: the sheer size of the teaching profession; its domination by women; its segmentation into numerous noncommunicating public school districts; the current emphasis on individual treatment for the child; the economic problems of individual-institutional relationships; and, finally, the lack of previously determined criteria for excellence. On the other hand are other factors that would support an effective method of evaluation were it available: the existence of national professional organizations—many of which desire evaluation; the ideology of professional uplift; the need to improve classroom teaching; the need to protect teachers against unfair and arbitrary supervisors; and the growing public demand for identification of superior teaching. Thus, it would seem that teachers' reluctance to accept proposed methods of evaluation is a result (rather than a cause) of the present confused state. It is significant that this volume is commissioned by three national educational organizations—each interested in spurring effective research on the competence problem.

If none of these five explanations is tenable, why then has so little resulted from the millions of dollars spent on teacher competence research? What can be done to change this segmented and noncumulative picture to one of progress toward an important social problem?

The thesis in this chapter is that effective research on the competence problem cannot be made until a general theory of the factors in instruction has been accepted, and until programed, integrated, long-range research is organized around that theory. Much of the previous research on this problem is noncumulative because investigators have measured literally dozens of different types of phenomena using unique terms, methods, assumptions, and propositions. Until a broad, standard system of concepts is utilized, progress will not be made. The authors propose here that one such integrative system is the framework of role theory.

Parameters of the Problem

Before detailing the unique concerns, terms, and concepts of role theory, it is useful to examine the range of phenomena to be dealt with when studying teacher competence.

Antecedents, Teacher Behavior, and Consequences

Three persistent problems plague all attempts to define, demonstrate, train for, or measure teacher competence. These are: (1) level of abstraction, (2) time sequence, and (3) theories of the outcome. The first two are considered in this section, the last in the section following.

In dealing with a teacher, all that can be measured directly is overt. Observers will agree, for instance, that a teacher is tall or short, that he is speaking loudly or softly, or that attention is being paid (or not being paid) to Johnny. For convenience, reference will be made to teacher characteristics that are transitory, volitional, and understandable as *behaviors.* All other teacher characteristics (age, sex, skin color, physical states, shape of the ears) are *features,* and it should be noted that teachers have little or no control over them in most circumstances.

If teacher behaviors and features are overt, so are the physical and social characteristics of the immediate environment in which the teacher and others may be interacting. Thus, the teacher admonishes Johnny about running in the halls at 10:35 A.M. on a hot, sultry morning when no other persons are in sight. It is to be presumed that her behavior would change in various ways were she to observe Johnny running on the playground, in a crowded hall, at another time of day, and so on.

Having looked at the overt characteristics of the teacher and of the environment (including Johnny), what other aspects of the situation may be observed? *Nothing.* All of the really interesting aspects of the situation, both in terms of psychology and sociology, must be surmised from evaluating the event or from a pattern of similar social events. Thus, it can only be surmised that the teacher "perceives" the situation in a certain way, that she is "motivated" to do something about it, that the situation exhibits a pattern of "deviancy" similar to other patterns, that shared "norms" exist regarding appropriate behavior in the halls, and so on. None of these represents overt aspects of the situation.

Figure 1 is used to emphasize the dichotomy between overt and surmised aspects of the situation. In this figure, only the events between the dotted lines may be directly observed; characteristics of the teacher personality or of the social situation are surmised.

The arrows in Figure 1 indicate presumed cause-and-effect relationships. Thus, that teacher behavior is presumed to be a product of internal and external stimuli, and so forth. Teacher behaviors are the only overt characteristics of the teacher included in Figure 1 (since teacher features are unchanging and therefore relatively independent of both environment and teacher personality).

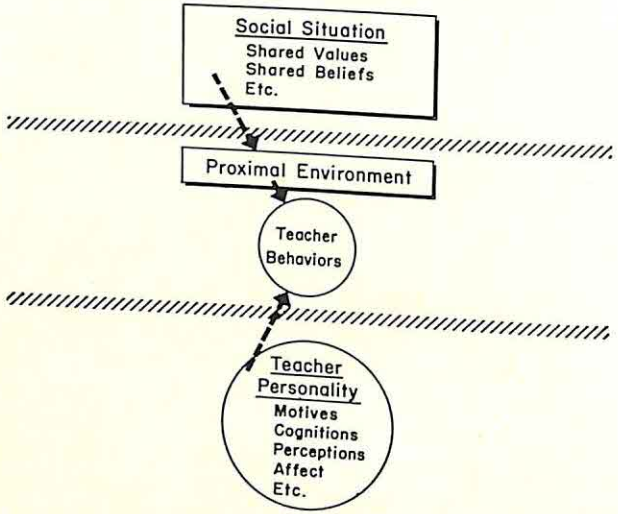


Fig. 1. Proximal variables involved in explaining teacher behavior.

However, in the little world of social interaction neither teacher behavior nor environment are truly stable. Johnny is impudent to the teacher when she corrects him, or he becomes reasonable and smiles. The principal tells the teacher that she is doing a good job, or he admonishes her for allowing too much noise in her classroom the previous morning. In such cases, it is reasonable to presume an interaction between teacher behavior and environment, a feedback model. Figure 2 shows such a model. Teacher behavior at time 1 influences the environment (in this case, pupil behavior) which, in turn, causes new facets of teacher behavior to be shown at time 2.

In Figure 2, changes in personality that accompany experience have been ignored. The personality of the teacher must be slightly different after experience in the proximal environment of time 1, and her behavior at time 2 reflects that change. Similarly, it is hoped that teacher classroom behavior had an effect upon pupil personality and produced more than merely superficial pupil conformity.

Figure 2 demonstrates something of teacher-pupil interaction. However, in the wider field of teacher competence evaluations, the teacher usually is placed in a larger time sequence. This sequence runs from her early childhood or educational experiences through her behavior as a teacher to her ultimate effect upon the lives and prospects of her pupils. This broad view is demonstrated in Figure 3.

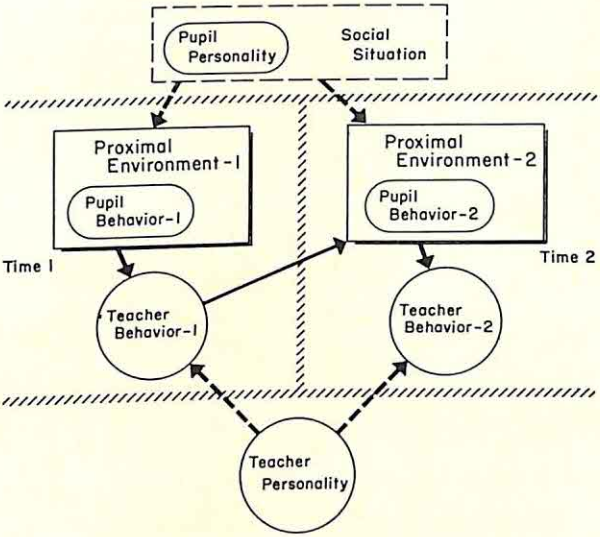


Fig. 2. Proximal variables involved in explaining teacher-pupil interaction.

The most obvious characteristics of the interaction situation (teacher behavior and the proximal environment) have dropped from view almost completely. Now under concern are a triumvirate of relationships involving formative factors, teacher personality characteristics, and effects of the educational process. The time span is broad, the abstractions are high order, and the realities of individual interaction are often ignored completely.

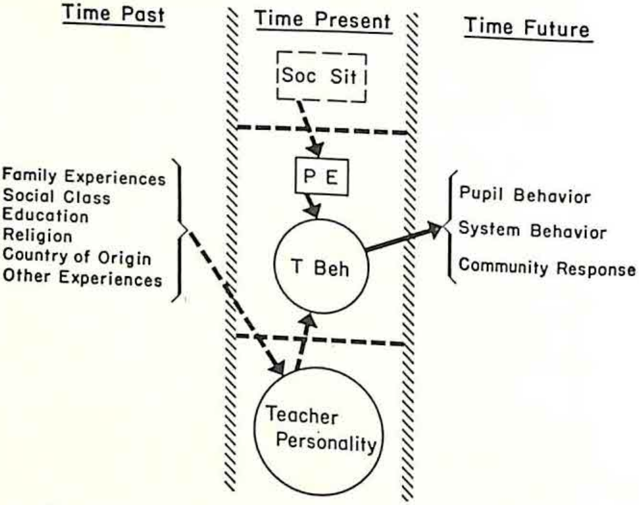


Fig. 3. Distal variables involved in explaining teacher competence.

It is useful in examining Figure 3 to consider three types of teacher competence definitions which often have been given.

Definition 1 Teacher competence is the ability of teachers to accomplish the (unspecified) goals of education, and it is measured best by examination of previous experience or of demonstrated level of achievement.

Operational measures based upon the first definition often are concerned with the number of years of formal education. Thus, teachers are given pay raises because they have survived a specified number of additional hours of graduate training. The difficulty with this definition is that it ignores many of the processes illustrated in Figure 3—the relationship of background, personality, and behavior to educational outcomes. Unfortunately, most empirical studies of academic achievement and subsequent competence (in fields where the latter is independently measurable) show low correlations.

> ***Definition 2*** Teacher competence is a characteristic of teacher personality that leads to achievement of some (usually unspecified) goals of education. This is best measured by "personality tests."

Although the second definition appears to have more merit than the first, it suffers from many of the same defects. First, it presumes a one-to-one correspondence between teacher personality and behavior. Second, it presumes that teacher behavior is independent of environment, or, even worse, that it can transcend a bad environment to produce good results for all pupils. Let us consider only the outcome of pupil learning. It is probable that this variable depends upon reading materials available, the presence (or absence) of machine aids, the social climate of the school, the social position of the child in his class, the intelligence of the child, and many other environmental aspects that may interact favorably with the teachers' best efforts or reduce their effect.

> ***Definition 3*** Teacher competence is teacher behavior that achieves a given educational goal.

This definition is less useful than either of the two preceding it. It does not suggest the means by which one may train for or produce the relevant educational goal. Also, it allows for uncontrolled byproducts of teacher behavior. However, many school systems have attempted to measure teacher competence simply by averaging the number of months in which a rise in pupil achievement is recorded during the year. This practice is not recommended.

Judges and Judgments

All definitions of teacher competence have a point in common —teachers are to be judged on the basis of their ability to produce certain outcomes or goals. The goal concept will be examined in some detail.

It is possible to define an outcome in neutral terms. Observing a man sawing a board, one concludes that, if he continues to saw in the same manner, he soon will have it cut through. Teacher behavior may be observed to be progressing toward certain outcomes—teacher-pupil interaction may be said to be becoming more stable, more learned, and so forth. However, simple *identification* of outcome is inadequate for competence evaluation. Few persons would judge as competent a teacher who produced only riots or who encouraged immorality in her pupils, no matter how effective she might be in bringing about such actions.

A goal, as opposed to an outcome, involves the attachment of affect—of approval or disapproval by one or more persons—to an outcome. Goals may be individual ("I want some of that good peanut butter") or shared ("We should not get into a nuclear war"), but they must always involve at least one person. Thus, competence of the teacher, inevitably, must be judged against the standards of people who may or may not agree with one another over the desired outcomes of teacher behavior.

Before competence can be judged, then, an agreed-upon set of goals must first be established for the educational system (at least in the local community). If there are no agreed-upon goals, one man's competent teacher may be another man's ineffective nincompoop.

The problem becomes more complex because goals may be short or long range. In fact, for some persons, teacher activities themselves may have a goallike quality. Thus, for some persons, "corporal punishment" is to be avoided in the school regardless of the situation—even though its use might encourage learning in a particular pupil. On the other hand, more sophisticated persons are apt to judge teacher competence in terms of pupil behavior expressed, perhaps, many years in the future.

The problem is also complex because of the varieties of outcomes that may result from teacher behavior. Focusing upon the individual pupil alone, educators commonly measure academic achievement, attitudes toward education (and many other things), social skills, creativity, and physical health and prowess. However, outcomes of teacher behavior are not confined to the individual pupil. Many teachers complain that they are judged by the principal in terms of how quiet they keep their classrooms or how disciplined their group is on public occasions. Teachers with group-dynamics training can become depressed when they note an increase in the cleavages or number of social isolates in the classroom.

Apart from these criteria (which are intrinsic to the purposes of education) teachers also are judged as members of an educational team, on how willing they are to sell tickets or coach basketball, on their participation in community activities or attendance at PTA, on their personal habits (even in their homes), or on their extracurricular activities. Apart from whether or not such judgments are legitimate, teacher competence can also be judged in terms of "public image" or "community service," and so on.

Definition, Measurement, and Development of Competence

The complexities of the educational process and the ineffectiveness of competence research to date both suggest the following definition:

> ***Teacher Competence*** The ability of a teacher to behave in specified ways within a social situation in order to produce empirically demonstrated effects approved by those in the environment in which he functions.

This definition suggests the exploration of a series of empirical and normative questions before teacher competence is evaluated. Figure 4 suggests some of these questions.

The outcomes shown in Figure 4 also encourage pessimism in regard to the immediate resolution of questions involving competence. Not only is an adequate theoretical model lacking to handle this problem, but also missing are specific empirical information that should precede research on competence itself (such as a classification of social situations in the classroom). Not only is information lacking on whether judges are in agreement on the outcomes of teacher behavior, but many in our society would argue against such agreement.

There may be even more pessimism about achieving an effective "competence creation" or evaluation. If there is no definition for those behaviors in context that identify the competent teacher, how can one explore the background factors that lead to these behaviors or develop instruments that will measure ability to perform them? Answers to such questions are not readily available.

Such pessimism should not lead to inactivity, however. Figure 4 also suggests that progress will be made toward solving the problems of teacher competence by undertaking an integrated program of research on several fronts using a standard set of terms and concepts. Below the authors present role theory as one component of such a set. Following this is a report on some pertinent findings from a study of teacher role made in Kansas City. At the end of the chapter the authors return to the means by which role theory may be applied to the questions in Figure 4.

Role as a Theoretical Approach

Role theory may be said to deal with patterns of behaviors (or other characteristics) common to persons. Role theory also deals with a variety of cognitions held about those behaviors by social participants. The conceptual distinctions of role theory center around a description of the behaviors and cognitions. The propositions of role theory are concerned with the effects of the behaviors upon the cognitions or the effects of the latter upon the former.

Role theory has become increasingly popular in the last decade. Not only has it appeared in theoretical and empirical studies in social psychology, sociology, and various other social sciences, but it has also been adopted by group dynamics practitioners, adult educators, mental health proponents, and others interested in the practical applications of social sciences. However, it must be pointed out that role theory today is not precise and is not based on an agreed-upon set of "facts" or propositions. As has been pointed out by Neiman and Hughes (1951), Gross, *et al.* (1958), Biddle (1961), and other reviewers, role theorists are at odds with one another over what they are studying, what to call it, and what they know about it. Role theory today more closely resembles an area of interest—expressed in a more or less common language—than it does an integrated field.

Despite this, role theorists have tended to deal with only a limited number of concepts in their attempts to describe and account for human behavior in social contexts. The next two

Empirical questions	*Normative questions*
1. To what extent are judges in agreement on goals for judging teacher effectiveness? What goals are suggested by national tradition, local tradition, educational theory, and the law? How may conflicting standards be reconciled?	
	2. What areas of outcome should be adopted for use in judging teacher effectiveness? In the event of conflict between judges (the public, professors, lawyers, and so forth), who is to determine outcomes?
3. Can relationships be discovered among teacher behavior, social situations, and educational outcome? Can a classification be found for social situations (in and out of the classroom) that make possible the isolation of effective teacher behaviors?	
	4. What areas of teacher behavior are to be considered acceptable as ends in themselves? What types of social situations, techniques, devices are to be incorporated into the educational process?

Fig. 4. Questions that must be answered in order to develop teacher competence.

headings cover those distinctions expressed in terms found most useful by the authors.

Basic Distinctions

In general, five types of concepts are necessary in role theory. These are characteristics, types of cognition, levels of cognition, person distinctions, and backgrounds.

CHARACTERISTICS On the overt level, human beings exhibit only two types of characteristics—behaviors and features. Both characteristics have been discussed earlier. Behaviors are assumed to be overt, transitory, patterned, voluntary, purposeful, meaningful, or directed. Behaviors may also exhibit patterns that depend upon context (features persist and are independent of context).

Although behaviors *are* patterned, transitory, and voluntary, untrained observers do not necessarily operate with this in mind when dealing with behaviors. Frequently, certain behaviors may be viewed as if they were as persistent as features. These may be referred to in everyday language—crankiness, stupidity, docility, and so on. Often there are stereotyped references involving "frozen" patterns of behavior—for example, a minority group may be called "stupid and unclean." Thus, on the covert level allowance is made not only for overt characteristics, behaviors, and features, but also for the cognitive mapping of "frozen" patterns of behavior—traits.

TYPES OF COGNITION Role theory is usually concerned with two orientations to interaction—descriptions and prescriptions. The descriptive orientation is concerned with subjective probability or assessment of the world as it appears to the person. A descriptive orientation applied to the behavior of another is called an expectation. This may be defined as a cognition consisting of a belief held for a characteristic of a person or position. The prescriptive orientation deals with value, with statements of "ought to do," or with right or wrong. A prescriptive orientation applied to the behavior of another is called a "norm." This may be defined as a cognition consisting of an approval (or disapproval) judgment held for a characteristic of a person or position. Expectations tell us how the respondent views the world: "I think teachers generally are punctual." Norms tell us about his values: "I would like it if teachers weren't quite so punctual." Norms and expectations are conceptually independent, although obviously they may be related in many situations.

Both the norm and the expectation actually are two-dimensional figures. If communication about the same content area is assumed, expectations and norms may be displayed as in Figure 5.

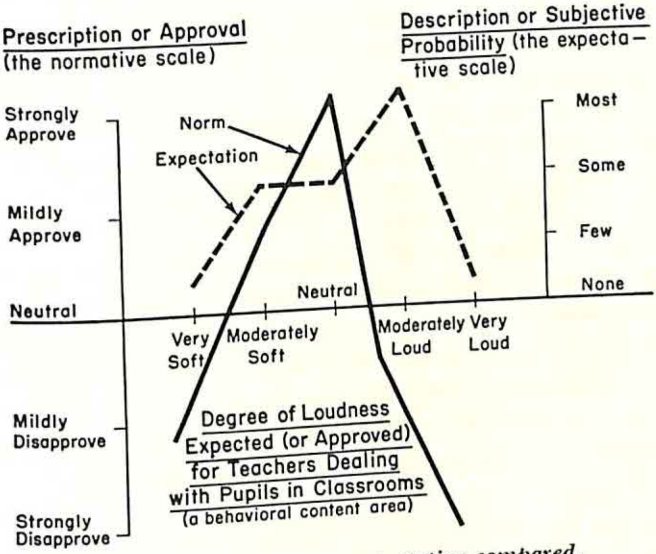


Fig. 5. A norm and an expectation compared.

It should be observed in Figure 5 that, whereas expectations vary only in absolute value, norms have a two-signed orientation running from negative to positive values. This distinction in sign has fostered the popular terms, "positive norm" and "negative norm" (referring to the points of maximum approval and disapproval). In the example given in Figure 5, at the behavioral alternative "moderately loud" the respondent both mildly disapproves of this activity and expects that it will be shown often by the person or position discussed. In such a case, a disparity exists between norm and expectation.

In addition to prescription and description, other cognitive orientations undoubtedly have value for predicting human behavior—power, legitimacy, and others. However, only norms and expectations have been widely studied to date in role research.

LEVELS OF COGNITION It has been pointed out that persons cannot operate in social relationships without estimating or appraising the cognitions of others. This leads to a recognition of two levels of cognition in role theory. "First-order" or "own cognitions" are those held by an individual and have as their focus the overt characteristics of another. Own cognitions include norms and expectations and may be held about behaviors, features, or traits. "Second-order cognitions" are those imputed by an individual to a second person in order to assimilate the overt characteristics of a third. Second-order cognitions also may consist of either descriptive or prescriptive orientations. This leads to a new family of terms. "Attributed expectations" are beliefs held about the expectations of others. "Attributed norms" are beliefs about the norms of others. "Ausvalued[1] expectations" are values held about the expectations of others, while "ausvalued norms" are values held about the norms of others.

Several distinctions should be made between first- and second-order cognitions. (1) First-order cognitions may be judged against overt events and thus checked for truthfulness. Second-order cognitions may be judged against the reports of others (a check termed "accuracy"). (2) First-order cognitions may involve only two persons—the holder of the cognition and the other(s) about whom the cognition is held. Second-order cognitions may involve three persons—the holder, the other(s) about whom the cognition is held, and the third person(s) to whom the cognition is attributed or ausvalued. (3) The existence of a second-order cognition does not imply that it is an accurate reflection of the first-order cognition held by the other, any more than a first-order cognition implies a truthful mapping of person characteristics.

PERSON DISTINCTIONS A distinction is made in role theory between persons and positions. Position (or status) has been used in several ways in the social sciences, but the word is most commonly used to point toward a place (or "slot") in the social structure. A "position" is a place in a social system bounded by rights and duties. Village chief, sister, corporation executive, and public school teacher are examples of positional designations found within a system of social relationships. For our purposes, "position" may be considered to involve a set of persons who exhibit similar characteristics, who are treated similarly by others, or for whom a cluster of unique cognitions are maintained either by others or by the persons concerned. Positions can involve many persons, few persons, or (in rare instances) no persons at all.

---

[1] The authors apologize for the neologism but they have yet to find an English language word with adequate connotations.

It is convenient to adopt a standard set of terms for use in referring to the persons (or positions) involved in role propositions. The literature abounds with person distinctions such as "respondent," "judge," "focal position incumbent," and so on. For convenience's sake, the authors have developed the following five persons into a symbolic interaction model: the "subject" is one who holds cognitions about behavior; the "object" is one who behaves (and about whom cognitions are held); the "target" is one toward whom behavior is directed by the object; the "audience" is one who may observe behavior but who is not a target; the "sentient object" is one to whom there is imputed a second-order cognition by a subject. It should be understood that use of this five-person model in no way implies that two or more of these persons may not coincide. A person may hold or attribute to others norms for himself, for example, or a subject may attribute *to* teachers expectations *for* teachers.

BACKGROUNDS Both behaviors and role cognitions ordinarily are set in a context or background. The fact that specific settings limit and modify the behavior of the person may be observed by watching him in different situations—at church, at the grocery store, at a night club. Barker and Wright (1954) have demonstrated the power of backgrounds to alter the patterns of behavior, while Biddle, Rosencranz, and Rankin (1961, Vol. 2) have demonstrated a similar effect in role cognitions.

The authors have found several background distinctions useful in role theory. A "setting" is a physical location consisting of an environment (and its manipulatable contents) in which behavior can take place. As used here, the term "setting" refers only to physical qualities of the background and not to persons present or patterns of behavior. Two other terms are necessary. A "position set" is a group of positions present in a setting. Each is represented by one or more members. A "social situation" is a pattern of behavior tending to involve all persons in a setting who interact in a characteristic, understandable fashion.

One last background, the "purpose," has been more of a problem for role theorists than any other concept. As already observed, one of the qualities of behavior is its tendency toward outcomes or "end states." Occasionally, (particularly when the other backgrounds are vague or the positional identification of the object person is not known) the context is stated as an outcome or a purpose. Thus, a purpose may be considered an outcome toward which the behavior of the object is presumably inclined.

There are several difficulties involved in the use of purposive backgrounds. They tend to be logically circular because they are based on an expectation presumed to be shared by all. To the subject, they almost always imply motivation (whereas the behaviors on which they are based may be accidental or a function of other background characteristics). They often suggest definitions that specify nothing about teacher behavior and are reminiscent of the third poor definition of teacher competence offered above. For example: "The role of the teacher is to instill knowledge into the minds of pupils." Finally, such backgrounds often are termed "functions"—one of the most misused words in the social sciences.

Once again, it is quite possible for backgrounds other than the four discussed here to be used in role theory. However, these four have received the most attention to date.

Derived Concepts

A number of concepts have been used in comparisons of role cognitions with one another or with their behavioral or first-order models. Even less standardization is found in the literature regarding these comparison terms. In general, however, concepts of comparison have involved assumptions that: (1) each element compared was appropriate to a given object person or position;

(2) each element involved an identical content area; (3) at most, a single cognitive orientation was involved.

A partial list of useful terms should include the following: "uniformity"—two identically defined role elements are said to be uniform if they are similar; "conformity"—a behavior is said to conform to a role cognition if it is similar to that cognition; "veridicality"—a cognition is said to be veridical if it is similar to the object characteristic it purports to map; "accurate"—a second-order cognition is accurate if it is similar to the sentient object's own cognition it purports to map; "shared cognition"—two or more subjects are said to share cognitions if those cognitions are uniform; "consensus"—two or more subjects reflect consensus when shared role cognitions are recognized, when subjects accurately attribute to others cognitions they also hold; "attributed consensus" —a subject is said to attribute consensus to a sentient object if he attributes to the others cognitions that correspond with his own.

In the past several years a number of authors have pointed out that role-element disparities may be disturbing to either the social system or the individual. Two terms frequently are used in these discussions—"cognitive dissonance" and "role conflict." The first of these relates to problems for the subject, the second to problems for the object. Cognitive dissonance occurs if two incompatible cognitions are held by the same subject. Cognitive dissonance is significant, according to Festinger (1957), because of its tension-producing effect on the individual. This, in turn, motivates further behavior of various sorts. Role conflict occurs if two or more incompatible cognitions apply to a common object. It is not required that the cognitions involved in a role conflict be held by the same subject. (However, if the object person is to be affected by role conflict, he must either become aware of the incompatibility or must be imposed upon by its behavioral outcomes in the actions of subjects.) The distinction made between cognitive dissonance and role conflict does not necessarily mean that these are independent processes.

Finally, a number of authors have pointed out an area of significance for role theory with the term "sanctioning." As used here, a sanction refers to a behavior (or a cognition about that behavior) that is directed by a subject person toward an object in order to get the object person to conform to cognitions held by the subject. The subject judges the degree of conformity of the object's behavior to his own expectation (or norm, and so on) and then "rewards" or "punishes" the object accordingly. It should be pointed out that, in all probability, the object person also holds role cognitions about this sequence of behaviors (as may others in the situation) and may not "notice" that the subject is sanctioning him. On the overt level, sanctioning is said to occur if behaviors are noted to be in a feedback relationship and if the "controlling" pattern of behavior appears to have gratification potential. On the covert level, sanctioning may be thought to occur only if the full feedback sequence is identified by the subject or object.

It should not be assumed that this short review is intended to cover the wealth of basic and derived terms that have been stated and examined in the role literature. Rather, concepts have been chosen to make possible a discussion of the authors' study of teacher role in Kansas City and the problems of applying role theory to teacher competence evaluation. The interested reader is referred to Biddle (1961) for a fuller discussion of these and other concepts.

The Kansas City Teacher Role Studies

In July 1958, a program of research on public school teacher role was inaugurated by the authors and others at the University of Kansas City.[2] This program had three broad objectives:

1. To pursue role theory and aid in its development as a sophisticated system of analysis in social psychology.
2. To study extensively and empirically the role of the public school teacher.

---

[2] This research was supported by the Cooperative Research Program of the Office of Education, Department of Health, Education and Welfare through research contracts granted to the University of Kansas City and the University of Missouri (to which the project moved in September 1960). Additional support for work in role theory has been provided by the Office of Naval Research (Group Psychology Branch).

3. To explore the relationships between teacher role and the problems of teacher recruitment.

These objectives have been pursued on a broad front. The authors have conducted systematic reviews of various areas in the role literature. These have included empirical studies relating to educational roles (Biddle, Rosencranz, and Rankin, 1961, Vols. B and D). They also have designed methodologies specifically for the study of role (Biddle and Simpson, 1961). Finally, a series was conducted of three empirical field studies on public school teacher role in the Kansas City metropolitan area (Biddle, Rosencranz, and Rankin, 1961, Vols. 1 through 5).

The first of these three studies, the pilot study, utilized a short questionnaire that was presented to approximately one hundred college students. This study was designed to pretest measuring instruments and to determine whether or not clear patterns would materialize distinguishing own from attributed cognitions of public school teacher role. Results from this study are reported in Rankin, *et al.* (1960).

The second, the background study, dealt with the contexts of teacher role cognitions. Approximately 225 respondents—representing eight different positions related to the public school—answered a questionnaire requiring more than two hours to administer. Backgrounds studied included the settings of both public school and community, social situations in school settings, and the purposes of teacher activity in community settings. From this study, a series of backgrounds was isolated for further investigation.

The third, the role study, attempted to identify the content areas, representing teacher characteristics, used by respondents in dealing with teacher role and to study the own and attributed cognitions held by respondents for teachers in those content areas. Approximately 1000 respondents representing ten positions were interviewed. These included teachers, parents, pupils, school administrators, and college students. Data were gathered concerning teacher behaviors, traits, problems, and motivations. The relationship between expectative and normative orientations was studied. In addition, a limited attempt was made to gather data on actual teacher behaviors by asking teachers to report their own activities.

Samples Used in the Role Study

Methods were similar in the background and role studies, but only those for the latter will be described in detail. In this study, a sample of fifty representative schools was chosen from the 406 individual public schools to be found in the greater Kansas City area. Using a table of random numbers, a proportionate number of schools was chosen for each of four community types—urban, suburban, small city, and rural communities. These schools, in turn, were matched against several other population characteristics to insure the sample's representativeness.

Respondents in the study included 98 teachers, 261 parents, 237 pupils, 22 school supervisors, 20 school board members, 23 principals, and 22 school district superintendents. Each of these persons, representing community positions, was chosen from one of the sample schools or its school district. Numbers of respondents were chosen so as to be proportionate to the number of position members associated with each school (however, school administrators were deliberately oversampled). Respondents in the study constituted 68 percent of persons invited to participate. It is likely that the final sample was not fully representative of ethnic minorities, low economic classes, professionally busy persons, and individuals with little interest in education.

Although the final sample was not a demographic sample of the public in Kansas City, it was representative of seven positions oriented toward the school system. Indeed, a truly "representative" sample of respondents might not be useful in studying teacher role, for only certain persons actually participate in decision making on public education issues. These were the people chosen for the sample.

Three groups of undergraduate students from the University of Kansas City also were included in the role study. There were 109 education majors, 79 students with majors in other fields, and 56 students who had not chosen a major at the time of the study. The student body of this university is somewhat older than the average, more mature, more often self-employed, middle class, and apt to be living at home. The university maintains a close relationship with the school districts of the region, and many education students are placed as teachers in the area. It is reasonable, therefore, to compare University of Kansas City students' role cognitions with those of community members from the area—particularly teachers.

Instruments of the Role Study

Several different forms were administered to respondents in the role study during a two-and-one-half hour group interview session. Three of the forms will be described here. Form One consisted of a set of sheets upon which respondents were asked to give their positive and negative norms and their expectations for teacher behaviors in specified backgrounds. Each sheet was headed by one item from a list of sixty contexts to be found in the public school (all of which had previously been isolated in the background study). Respondents were asked to indicate what they most wanted, least wanted, and generally expected teachers to be doing in such a background. Answers given were open-ended. Each respondent received eight separate background sheets, and up to twenty answers were given on each sheet. Data from this form were coded using a five-dimensional code designed to express the contents volunteered for teacher behavior. The dimensions of this code included activity, manner, imputed cognitive structure (such as teacher motivations or other high-order abstractions used by respondents), social target, and purpose. Within each dimension, a large number of content categories was provided. This resulted in over one million possible content combinations of unique codes. Altogether, over 50,000 responses were coded and then punched onto IBM cards for data processing.

Form Two was similar to Form One in that it consisted of a sheet upon which respondents were asked to volunteer their positive and negative norms and expectations. However, in this latter form, no background was specified. Respondents were asked to give cognitions for teacher traits. A modified form of the code used for Form One was also used in coding responses to Form Two. Approximately 10,000 responses were coded and punched from Form Two.

Form Four was designed to obtain comparable respondent data dealing with own cognitions for teachers and cognitions that respondents attributed to three groups—people in general, teachers, and school officials. It differed from Forms One and Two in several ways. First of all, a series of fifty scaled teacher behaviors-in-context (drawn from responses to Form One) were specified to respondents. Second, respondents were asked to check one of five possible behavioral alternatives under each question ranging from "Teachers would do a great deal" of the behavior involved to "Teachers would do little or none." Finally, respondents gave a total of twelve answers for each teacher behavioral item. They gave their own expectation and attributed expectations to each of the three object positions; then they gave their own positive and negative norms to each of the three positions. (Teacher respondents also were asked to check their own behavior for each item on a separate page.) In order to minimize respondent fatigue, Form Four was divided into five parts so that each respondent dealt with only ten behavioral items. The precoded responses given to Form Four also were punched onto IBM cards for analysis.

Several other forms were used in the role study in addition to the three described here. Additional forms dealt with problems faced by the teacher and with background of respondents. Three separate instruments were designed to elicit information about motivations to teach. Other instruments were used in the background and pilot studies. Altogether, over 1000 questions, items, or scale forms were designed and used in the three studies.

Methods of Analysis

Extensive use was made of IBM machines and computers in processing the data from the three studies. In general, these analyses followed a standard procedure.

1. CODE DEVELOPMENT Numerous codes were developed to express differences in content or degree of role cognitions held by respondents (and other information). In some codes, developed

on a priori grounds, respondents were asked to choose among forced-choice alternative answers to a given question. Other codes were developed after responses were volunteered by respondents. This type of code was necessary in areas where, to the researchers' knowledge, no previous investigations had been reported. Thus, there was nothing to lean on in preparing forced-choice alternatives. Postcoding was done primarily in areas of role content—content of background, behavior, trait, or feature in role cognitions held for the teacher. All codes were designed to reflect the data given, and all were refined so as to give a minimum reproducibility of 90 percent identical assignment of data to code categories by independent coders.

2. CODING All data then were prepared for machine tabulation by coding responses and editing questionnaires. These steps were performed by trained coders whose work was checked regularly.
3. PUNCHING AND VERIFICATION Data were punched onto IBM cards and were verified immediately. Generally, one card was punched for each response—resulting in a large number of cards for each respondent.
4. TABULATION Data were then prepared in tables showing relationships between independent and dependent variables. Where appropriate, instructions were given to tabulating equipment to insure that duplicate responses from a given respondent were ignored. All tabulation was done automatically. Tabulated results then were scanned for significant results.
5. COMPUTATION OF SIGNIFICANCE Selected results then were given to a computer for computation of one or more statistics representing types of comparisons to be found within the data. Entry of data into the computer was accomplished by using either the original data cards or secondary cards that represented tabulated results (and that were prepared automatically for the purpose).
6. WRITE-UP Finally, results were interpreted and written analyses were prepared.

Limitations of the Studies

The findings of the Kansas City studies have limited applicability. First of all, there were limits imposed by the conceptualization that dealt with role theory. The reader already has noted that, in defining "role" in cognitive terms, the authors have ignored several major issues that appear in the competence evaluation question. Other conceptual limitations are involved in the assumption that role cognitions are held by individuals (rather than by social groups in consensus), that orientations are limited to descriptions and prescriptions, that cognition types involve only first- and second-order cognitions, and so forth.

Second, numerous limitations in the design should be noted. A "funneling" design was attempted in which respondents would indicate contexts, role content, and finally exact role cognitions for the public school teacher. This funneling design was incomplete, and short cuts are apparent. For instance, position-set backgrounds were ignored. No study of actual teacher behaviors was made. There was no adequate study of nonschool settings.

Finally, the sample of respondents was limited in several ways. Not only were samples not demographically representative, but the entire study was conducted in a single Midwestern, urban center. Thus, it is questionable that some of the results would apply to the isolated small town or to an urban center on the East or West Coast.

On the other hand, nothing like the Kansas City teacher role study has previously appeared in the literature; many of the findings from this study were unexpected and provocative, and some of them may be applicable on a national basis.

Selected Findings for Teacher Role

More than one hundred independent findings have been reported for the Kansas City teacher role studies, and it would be impractical to attempt their summary here. Several should be of interest in connection with teacher competence evaluation.

1. THERE EXIST BROAD TEACHER ROLE STEREOTYPES THAT ARE SHARED BY NEARLY EVERYONE Two kinds of stereotypes predominated in the findings. The first related to the "indefinite extension" of teacher role into almost all kinds of moral issues. In both positive norms and expectations volunteered for the teacher, respondents tend to picture the ideal teacher as Jack Armstrong. Teachers are (or should be) courteous, kind, obedient, cheerful, and so on. Furthermore, teachers are "mousy" in that they do *not* habitually violate these approved behavioral standards. On the other hand, respondents do not share stereo-

> types about unwanted teacher behaviors. This finding tends to support the position that teachers live in a "goldfish bowl," and that Americans are not willing to allow a teacher to relax and leave the job—ever.

In contrast, a second and even stronger stereotype may be found among respondents who indicate that "the public" is much more conservative in norms for the teacher than are respondents themselves. As far as respondents are concerned (teachers, parents, school officials alike), the public wants teachers to crack down on pupils, to avoid smoking and drinking, to conform to the dictates of school officials, and so on. It is significant that, as far as the data are concerned, this is a *false image*. Not one of our groups rates itself as conservative in its own norms as it rates people in general. Moreover, respondents are alike in attributing similar conservative values to school officials (who actually are informed and liberal respondents). Such a false stereotype places severe restrictions on the behavior of school officials who must act to "represent" a consensus that, in fact, does not exist and that is clearly in violation of their own standards.

2. DESPITE THESE BROAD STEREOTYPES, THERE ARE SIGNIFICANT DIFFERENCES AMONG ALL RESPONDENT POSITIONS REGARDING TEACHER ROLE Moreover, the differences among respondent positions are not small. For example, parents, teachers, and school officials are seriously at odds with one another over matters of attendance and performance at PTA and other similar meetings. Parents see teachers as desiring to avoid PTA, while teachers, in fact, wish to attend and do attend. School officials see teachers as wanting to speak out (so do parents), while teachers attribute to these same groups a desire for teachers to keep quiet. In fact, teachers say they do not speak out—this in violation of their own norms. Other major areas of contention are matters of discipline, techniques for instruction, teacher self-indulgence, and relations with others in the school system.

3. THERE ALSO ARE BROAD DIFFERENCES AMONG TYPES OF COMMUNITY IN ROLES HELD FOR THE TEACHER Simplifying some rather complex findings, it has been found that small communities hold stricter norms for the teacher but are better able than large cities to predict the actual behaviors of teachers. In the urban center, the teacher is apt to be anonymous—to live miles from her place of work. A measure of concern is shown for out-of-school teacher behaviors in the community where a teacher re-

mains "a teacher" after hours. This gives respondents an insight into the realities of teacher behavior. (It is possible also that historical changes in teacher role are leading away from the small town or community evaluation of teachers and are emancipating teachers through a more limited definition of teacher role.)

4. SIGNIFICANT DIFFERENCES APPEAR ALSO AMONG EDUCATION STUDENTS, STUDENTS IN OTHER FIELDS, AND TEACHERS REGARDING ROLES HELD FOR THE TEACHER The greatest differences exist between teachers and noneducation students. This gives rise to the hypothesis that education students are undergoing an effective process of socialization in preparing for their profession. However, two other points are to be noted. (1) The largest differences in the responses of noneducation students and others involved expectations. Noneducation students felt that teachers were uncooperative, nonparticipating, self-indulgent, noisy, and restrictive with pupils. (2) Education students held autistic and idealistic norms concerning teacher-pupil relations. They would, for instance, allow greater pupil freedom than would either teachers or other students. It is not clear at present if these autistic norms result from, or tend to, stimulate recruitment into the teaching profession.

5. ACCURACY AND VERIDICALITY OF ROLE COGNITIONS ALSO SHOWED SIGNIFICANT DIFFERENCES DEPENDING UPON RESPONDENT POSITION Least accurate were those cognitions held by inexperienced respondents (such as pupils or college sophomores) or respondents at a relatively great social distance from teachers (such as parents or school board members). Older, more educated, more sophisticated, more centrally located respondents were better able to predict teacher behaviors and the norms and expectations of others.

6. FINALLY, ALL TYPES OF ROLE COGNITIONS, LEVELS OF COGNITION, BACKGROUNDS, CHARACTERISTICS, AND SENTIENT OBJECT POSITIONS SIGNIFICANTLY DIFFERENTIATED THE ROLE COGNITIONS HELD BY RESPONDENTS There were, for example, unique patterns for norms and expectations, for behavioral and trait cognitions, for own and attributed cognitions, and so on. Thus it seems clear that the role of the teacher is *at least* as complex as is implied by the conceptual system used, and that any additional work on teacher role should make these and other distinctions.

Application of These Findings to the Competence Question

Although findings from the Kansas City teacher role study were not developed out of a desire to evaluate teacher competence and are not directly applicable to competence issues as defined, they seem to suggest some important issues.

1. If a widely held, but mistaken, conception exists of a conservative "public" that holds restrictive norms for teacher behavior, decision making about education should be assumed to reflect this false image. Thus, teacher competence evaluation would tend to be resolved in terms of a false stereotype in a way that would please no one. Destruction of this concept would help to raise the attractiveness of the teaching profession, reduce role conflict for teachers, solve problems of public-versus-private behavior for school administrators, and allow competence evaluation to proceed along useful avenues.
2. If a widespread stereotype exists of teacher norms applicable to all issues of public morality, then not only can there never be a termination of the teacher's job, but it becomes coterminous with the Christian ethic. It would seem that benefits should accrue from a public discussion or educational campaign directed toward limiting the range of content area included in teacher role. This would, simultaneously, limit the grounds on which competence is to be judged.
3. If there are significant differences among respondents—by position and by type of community—in roles held for the teacher, can respondents be expected to agree about the outcomes of teacher behavior? It should be assumed that expectations for the teacher depend upon perceptions of teacher behavior, information gleaned from others, and the effects of neighboring cognitions (such as norms, cognitions attributed to others, and so forth). Norms will reflect these factors plus a subject's basic value structure. It is possible for a sophisticated respondent to hold expectations for teacher behavior that violate what he considers to be competent behavior patterns. However, the fact of normative disparities suggests, at the least, that respondents are in disagreement with one another over what constitutes a competent teacher. Such disagreements obviously must be resolved before the competence-evaluation question can be solved.
4. There were so many differences among respondents over accuracy and veridicality of teacher role cognitions that serious questions are raised concerning the ability of persons to serve

as objective instruments of competence evaluation. (In passing, teachers and school officials were most accurate in regard to cognitions held. School board members were no more accurate than parents.) Apart from their disagreements with one another over basic values against which to evaluate, the majority of our respondents (including most school personnel) were strikingly inaccurate and nonveridical in their role cognitions.

5. Finally, the existence of many different "pieces" of teacher role suggests that the competence issue should be laid in a similarly complex conceptual structure.

The Role Approach to Competence Evaluation

The Kansas City studies conceived of teacher role as a component (rather than a basic ingredient) of teacher competence. While an adequate study of teacher role should include the measurement of own and attributed norms and expectations about a variety of teacher characteristics, an effective study of teacher competence will come only when firm propositions are established to tie together patterns of teacher behavior and agreed-upon outcomes of the educational process. Nevertheless, a role approach seems to offer much in the solution of the competence problem.

Possible Contributions of the Role Approach

CONCEPTUALIZATION Role concepts can be expanded to contain findings of previous studies of teacher competence. With this as a starting point, future research can be undertaken with some chance of success.

TEACHER BEHAVIOR IN CONTEXT Role theory is designed to express the complex interactions between behavior and its physical and social contexts. Role theory offers an effective means by which teacher behavior may be described—particularly as it interacts with context. In addition, to the extent that cognitions about teacher behavior are used as methods of judging competence, role theory is an ideal vehicle for expressing these sources of judgment.

MEASUREMENT OF CONTENT As suggested in Figure 4, there are two areas of content involved in the competence question—those of educational outcomes and of teacher behavior. The role approach is precisely designed to deal with the latter. It may also be extended to the contents of educational outcomes.

THE RELATIONSHIP BETWEEN COGNITIONS AND BEHAVIOR Until now the focus has been upon the relationship between teacher behavior and matching role cognitions. There is another side to role theory—the relationship between role cognitions and subject behavior. For instance, it is reasonable to assume that when a parent finds that a teacher is seriously violating norms that the parent holds for teacher behavior, the parent will take action to bring about a change. Data supporting such a proposition have been advanced by Festinger (1957), Rommetveit (1954), and others. Role theory, therefore, is a good vehicle for examining the behavior of the community on issues of teacher competence evaluation. It should be particularly useful in helping to resolve the disparities in cognition (if not basic values) that must be managed before an agreed-upon method of evaluating competence is possible.

PERSONS AS JUDGES Finally, it seems enough evidence has already been accumulated from studies to indicate that untrained members of the educational community (parents, pupils, teachers, school principals, school board members, school supervisors, superintendents) are *not* likely to be objective judges of competence. In the first place, they differ widely in norms they hold or attribute to others (often inaccurately) for the teacher role. In the second place, many of them do not have much opportunity to observe the individual teacher in action. It remains to be seen whether or not a group of trained evaluators could be made an objective instrument for teacher competence evaluation.

Next Steps

If these contentions are correct, and role is an appropriate approach to the competence problem, several recommendations may be made for the next steps in empirical research.

1. First, a national study of teacher role is needed. It is possible that many of the findings from Kansas City will not hold or that others will take their place when teacher role is considered on a nation-wide basis. It is also possible that wide variations in regional treatment of teacher role will appear. To apply such a study to competence evaluation calls for a broad context of operations including questions pertaining to educational outcomes or goals. The study should also be contrasted with similar (or smaller) studies done in related societies—such as England or Canada. It should be emphasized, however, that much of the work of conceptualization and development of tools and methodology already has been done in the Kansas City studies.

2. Assuming the existence of a variety of regional differences in norms and expectations held for the teacher and the persistence of "false" images of public opinion, it would be useful to develop an instrument to enable the local school administrator to sample public opinion about teacher role. Such an instrument might be administered to a sample of respondents chosen by the administrator according to detailed instructions. Coding, processing, and evaluating dominant role elements found in the community would be done at a central facility and reported to the administrator on a standard form. Such a service would involve expense both in development and in administration but might be of great value in helping the local school administrator and school board make policy decisions.

3. Assuming that role conflict, false images, or conflicting values for outcomes of education are held by persons in the community, is it possible to eliminate these disparities by lawful democratic procedures? This question is a legitimate one that has been considered in connection with many other public issues (such as desegregation). Empirical research into such problems should examine the effects of educational campaigns, public discussions, *fiat* decisions by local school boards, and other methods in achieving a workable consensus. The authors feel that the lack of consensus in teacher role revealed in the Kansas City data results, in part, from the fact that these issues have not received a public airing, and respondents are speaking in a way that does not reflect carefully thought-out concepts.

4. Assuming at least local consensus on the role of the teacher, plus some reconciliation of contrasting values for education, and a reasonable reduction in the number of content areas of teacher behavior and outcome to be examined in the competence context, the next empirical step should be an examination of the relationships between teacher behavior and educational out-

comes. This would require a number of detailed studies comparing teacher behavior (in a social context—using the methods of behavioral observation) with dependent variables of both short- and long-range educational outcomes. A serious program in this field would take no less than 10 years as most of the important educational outcomes would appear only after a period of time.

In Conclusion

It does not seem likely that an adequate method for evaluating teacher competence will satisfy educators, social scientists, and the public. This pessimism is based on the conviction that, for many values related to education, disagreement persists among various groups of persons. For example, a university is often thought to have two functions—to preserve culture and to research into the unknown (which will result in cultural change). It has often been noted that while legislators will give millions of dollars for one, funds for the other may be much harder to come by. It seems that such value conflicts may well be irreconcilable. Perhaps the community will not be able to come to a public consensus on the aims of education, let alone recommend a course of teacher action leading toward those aims.

This, however, is not intended to minimize the usefulness of continuing to work toward at least partial solutions. Even if the entire problem cannot be solved, it is useful to discover what our values are and how to behave as teachers so as to bring about outcomes desired at the moment. Much educational research today may be understood to extend as far as how to produce a particular effect in Johnny without disturbing other aspects of the child's total potential.

Even if the ultimate goal—a workable means of evaluating teacher competence—is impossible, increased knowledge about teacher role, the aims of education, and their interrelationships will help to smooth the paths of educators. A role approach to these problems seems to be a useful one.

9

WARREN A. PETERSON

Age, Teacher's Role, and the Institutional Setting

From a study of the ways that teachers adjust to age and generational differences and to one another, some important facts about effectiveness have emerged. The teacher's role changes as she grows older; morale, and perhaps teaching competence, may be affected by age and generational conflict; and, most fundamentally, career teachers want recognition as they grow older—recognition for having dedicated their lives to other people's children.

The research reported here was based on comprehensive interviews with a relatively small sample of public high school teachers in a city of medium size and Midwestern location. The analysis techniques were those of comprehensive interviews with emphasis on the development of insights. Techniques were qualitative rather then quantitative and concerned social patterns and processes rather than prevalence or reliability. The validity of the data rested on the honesty and competence of teachers as observers of their own role in a social system. The objectivity of the results depended upon the responsibility and degree of insight possessed by the interpreter.

Excerpts from interviews are presented as fully as possible in order to document additional points and relationships found elsewhere in the chapter. In general, the processes described are assumed to be real—for there is no reason why teachers should have invented them. An understanding of the importance of these processes and their effect on other aspects of the teacher's role must come from further research.

Sampling and Interviewing

The present research differs from previous occupational studies of teachers in several respects. Although most of the analytical problems are not new, the range of inquiry was broadened to include the full work career with special emphasis on age and career changes. Thus the intent was to deal with a broad range of analytical problems—but a relatively small and homogeneous group of subjects.

The sample (confined to white, female, high school teachers thirty to seventy years of age) was drawn, at random, from an annual directory of personnel in an urban school system. In 1953, when the sample was drawn, less than 12 percent of the teachers in the system were under forty years of age and less than 7 percent were married. Sampling ratio categories for age and marital status were devised, and the sample was stratified accordingly. From an original sample of seventy-two teachers, fifty-six respondents completed the interview. The final sample closely approximated the original sample in distribution among high school units and among high school subject fields. Approximately two thirds of the teachers interviewed were single, and approximately two thirds were reared in small towns and rural areas.

The principle data-gathering technique was an informal, conversational interview. Sessions were structured partially by topics but were largely unstructured in the wording and the order of questions. Interviews were conducted in the respondents' homes and averaged two hours in length. Unless the respondent objected, interviews were recorded on tape.

The Sociological Frame of Reference

In recent years, an increasing interest has developed in the sociology of occupations and professions and in the closely related area of industrial organization. Attention is focused on these relationships because of a growing recognition that occupational groups constitute a basic form of differentiation in modern industrial society. Work is one of the principal ways that the members of an occupation relate to one another, acquire status, and participate in the social process. Further, the employed adult's work role is fundamental to his conception of himself as well as to his personal identity. As sociological attention has been directed toward the metabolism of social roles, occupations have been studied in an effort to determine how workers of one category may be expected to relate to those in other categories (colleagues, inferiors and superiors, clients, and outsiders). To the extent that the occupational role has integration, it emerges as a synthesis of adjustment or of different, and often contradictory, expectations.[1]

This set of relationships becomes more complex when viewed in career perspective. The problem in career research is that of discovering and explaining the role and status sequences that characterize members of an occupation during the span of life at work.

Aging is a basic element in movement through life—it is a common denominator. If aging is related to the varied conditions of modern industrial organization, it should also serve as a point of departure for generalizations about the metabolism of the occupational career. For example, among established middle-class occupations, there tend to be four distinct phases: (1) a learning-oriented-exploration phase in young adulthood accompanied by apprenticeship status and involving progressive acceptance or rejection by older, established members; (2) a secondary membership phase marked by high productivity in relation to the functions of the organization but lacking full recognition or responsibility;

---

[1] The analytical approach used in this research was developed primarily by Everett C. Hughes and his students. One of the earliest statements of this perspective may be found in Hughes (1937).

(3) a full membership phase that, in the middle class, seems to occur in middle or late-middle age and involves maximum formal and informal prestige, authority, and responsibility within the organization; (4) a period of membership decline characterized by deference rather than by prestige and by honorary, rather than functional, positions.

The development and validation of generalizations like the above are likely to occur through comparative research on several occupational careers. Although a general model has been employed as a guide in the analysis, there are more modest objectives in this chapter. The careers of a specific group of teachers are described and explained in terms of the structure of school systems, the character of student-teacher interaction, the pressures of woman-role expectations, and the effects of social class and rural-urban mobility. In addition to its implications in the direction of understanding common occupational career processes, this study helps to explain the special function of the teaching occupation in society and in the institutional system with which it is associated.

The Teacher in the School System

The teacher in the urban public school works in a social milieu that provides regular contact with pupils, other teachers, and principals. Less frequent contacts are made also with staff specialists from the central administrative office, with parents of pupils, and with service and clerical workers. Inducing socially valued changes in students appears to be central in the teacher's occupational purpose, her sense of mission, and her rationale for occupational commitment.[2] Within the administrative organization the teacher holds a staff-terminal position; but, because of professional status and the separation of classrooms, the teacher's role includes more autonomy and less direct supervision than that of many other staff-terminal workers. Terminal in the authority system, exposed to public opinion, subject to stereotyping, and set apart from many other occupations by the specialized occupational role (and by the predominantly female composition of its membership), teaching is characterized by intensive colleagueship and a pervasive occupational identity (Donovan, 1938).

---

[2] In summarizing his study of Chicago teachers, Howard S. Becker points out: "The teacher considers that she has done her job adequately when she has brought about an observable change in the children's skills and knowledge which she can attribute to her own efforts." (Becker, 1952, p. 453)

Deriving predominantly from the lower edge of the middle class and from aspiring farm and urban working-class families, teachers are engaged in social-class mobility as they enter the occupation and progress through the career (Warner, *et al.*, 1944; Whitney, 1927; Best, 1948). In many cases, the primary decision is to enter college—with teacher-training serving as the most available route for the academically or economically marginal (Valentine, 1950). At least among Chicago teachers, there occurs, in early teaching experience, a "reality shock" in coming to terms with problems of disciplining and motivating students—especially in lower-class schools—and in confronting the relatively hardened orientations of older teachers and administrators (Wagenschein, 1950). In this situation, as Becker (1953) reports, experienced teachers are concerned that young teachers may reduce the teacher group's authority by failing to maintain discipline or may change the standards of work performance through misplaced vigor and idealism.

As with most occupations in which women predominate, teaching is an "early plateau" occupation characterized by relatively minor changes after the beginning phases of training and apprenticeship. However, institutional arrangements are such that teaching offices are standardized from one unit of a city system to another and from that system to others that are politically and geographically separate from it. Occupational mobility among teachers, thus, takes several forms—movement from one grade level to another (most typically from grade school to high school teaching), movement to school systems of greater size and prestige with increases in salary and security, and movement from one unit of a city system to another. This latter kind of mobility has been analyzed in detail in Chicago where, until recently, the school system was constrained by an administrative policy that limited employment to graduates of its own teachers college. Under these conditions, with the individuals making up school units varying with the class and race composition of neighborhoods, a pattern of teacher mobility emerged from Negro and lower-class to white and middle-class schools.[3]

Career changes in the teacher's relationships with others of importance in the work setting have not been subject to any degree of empirical investigation. There are some indications that young adulthood is the most effective age for a teacher.[4] The Chicago studies, however, suggest that older teachers are more likely than young adults to be accepted into the colleague group, have greater power and authority in the school, and develop a greater interest in maintaining established procedures.

Relating to Students as Young Teachers

It seems important, in approaching the matter of relating to students, to point out the special age structure of the teacher-student group. This group is vastly different from the parent-child model, and the difference results from the application of the full adult, life-work career to teachers and the fragmentation of students into narrow age grades. A teacher begins teaching children when her age position to them is like that of an older sister; in midcareer she is about the age of their parents; and if she continues teaching to the age of retirement (age seventy in the urban system studied) she may be older than their grandparents.

---

[3] Comprehensive analyses of interschool mobility patterns among Chicago teachers are presented in Becker (1951) and Winget (1951).

[4] Waller (1935, p. 216) concludes: "The teacher-pupil relationship seems to realize its best possibilities when the teacher is a young adult, far enough away from adolescence to have solved his own problems but near enough to be understanding and tolerant, far enough from the students' world never to become a rival and close enough to maintain an interchange of feeling and ideas, far enough removed in age from his students to classify definitely as a member of the adult generation and young enough not to represent a horrible example of what normal young persons do not want to become."

In an essay showing great insight, Mead (1951, pp. 24-27) suggests that the teacher of classics to upper-class students in the old academy is close to the grandparental role; the teacher in the city slum school, reassuring and directing immigrant children toward assimilation and mobility, is close to the parental model; and the teacher in the modern middle-class school, "whose task is to stay close to the young child's body impulses and exuberant imaginative attempts to take in the world around him," is similar in function to the child nurse or the peasant nurse of other societies.

Except on occasions when they encountered students beneath them in class or ethnically strange (in mining, industrial, or cotton-belt communities) teachers interviewed gave a warm and glowing account of their feelings for students in early career. Although some of the "pleasant experience" feelings may have been due to the fact that some teachers were closer to their native surroundings than they were, later, in the city (or to the substantial recognition that teachers receive in small towns), many teachers attributed such feelings to age. In general, when asked about the best teaching period, either for themselves or for a teacher in general, they recognized a short orientation period, a "best teaching period," while still young and vigorous. Later, a decline was indicated.

I.: What do you think is the best period for a teacher? [5]

19s: Probably after one has had a year or two of experience, then while you are still young you have the most fun with the kids. Then later on—forty or so—it becomes more burdensome, after you get up in years.

In the small town the orientation period seems to be relatively short. For example, in discussing her first teaching experience, 16s moved rather rapidly from being terrified "the first hour" and insecure about her knowledge "the first year," to feeling that "it was a lot of fun."

16s: I know the first hour I was scared to death because I had students from a rural area—this was a small town—I had students almost as old as I was; big as I was, that's for sure. And I taught geometry and advanced algebra. That first year I was afraid they'd think I didn't know as much as they did about it. But it was a lot of fun, I really enjoyed it. I can't remember any particular problems.

---

[5] In the quotations given, "I" stands for the interviewer; respondents are identified with a number and a letter indicating marital status—"s" refers to single teachers, "m" to married, "w" to widowed.

In discussing early career experiences, teachers repeatedly stressed the idea that they were not much older than the pupils; this is an important aspect of "having a wonderful time." In 31w's case, discipline problems were approached with a sense of challenge.

31w: The first teaching job I had was in a small town. I had a wonderful time. Another girl and I went out there together. The pupils were just about as old as we were. We dated and had a wonderful time. It was supposed to be a difficult school. They had run out [sic] other teachers. But we gave them a run for it. We outsmarted them. I never had such a good time. I have friends there that I've had all these years. My partner married one of the boys. I had a good time with them; got more work out of them than we could possibly have gotten otherwise. It was an interesting town, hidebound in many ways. But, as I say, we gave them a run for it.

The young teacher in the small school usually applied herself with vigor and enthusiasm. In retrospect, there was bewilderment about how she was able to undertake so much. The close relationship with students was supplemented with warm relationships among young teachers, with joint activities with students, and with joint dating activities. There was, as 52m reported, "no distance between students and teachers."

52m: I taught English and did all sorts of other things. I sponsored a page on the schools in the local paper, sponsored the senior class, produced plays—I was the dramatics department, such as it was. And I started the pep club. I was quite busy.

I lived there with three other teachers who had rented a house. The other three had been there the year before, but they were young. The oldest one was 35, but she was very youthful and sort of the follower type. The others were just slightly older than I. It was very friendly, and we had a wonderful time. The kids would drop in all the time, whenever they felt like it. I think it was partly because they could get away with some things in our house that they couldn't at home. There was no distance between students and teachers. I have kept in contact with some of the students.

I did all sorts of things that year. I don't know now how I did it all. I produced "The Other Wise Man." How I managed to do that or how it turned out, I can't imagine. The other girls would stand on their heads to help in anything we undertook.

Perhaps because most teachers were older when they entered the city system, because of more severe beginner-acceptance problems there, or because of greater distance between students and teachers in urban areas, such special satisfactions from teaching were not generally emphasized in connection with first teaching. There were, however, a few references of this kind. 37s, for example, pointed to her energy and vitality when she first began teaching in the city at the age of thirty.

> *37s:* I had worlds of energy, and they had worlds of vitality, and together there wasn't anything that we didn't try. It has been a little difficult sometimes to keep from wondering why I can't do as well as I did with that first bunch. Maybe I am and don't realize it.

Age Changes in Relating to Students

Students are blamed for making the job more difficult as teachers grow older. One might say that students actively engage in age-grading teachers. When a teacher becomes informal with students, they are apt to question her about her age. Very significantly, this is associated with questions about why she hasn't married. A fifty-year-old unmarried teacher commented:

> *40s:* They are intensely interested in their teacher's age. If they could ask me one question and find out the facts, I suppose the first question that they would ask me is why I haven't married, and the second question would be, "How old are you?" Those are the two questions in which they are the most interested. They would not ask me anything about history; they would not ask whether I could cook or sew or if I make my own clothes or anything else, but "How old are you?" and "Why haven't you married?" I know because they ask me all the time.

In this instance, the teacher showed concern because students asked about the least presentable aspects of her personal identity—the old-maid features.

Student interest in the age of teachers is, apparently, one of the important factors in teacher sensitivity about age. This sensitivity was evident throughout the interviewing—as well as at the point when the interviewer attempted to secure the teacher's age. In the following quotation, a married teacher, in her early thirties, explicitly attributed this sensitivity to student interest.

I.: Why are teachers so sensitive about their ages?

52m: I don't know exactly, but they certainly are. I think it is partly because the kids are concerned about it. I have begun to notice it about myself, being sensitive that is. I have forced myself to tell them my age when they ask. I catch myself dodging and then grit my teeth and tell them. I have often wondered how they look on me. I think they consider me ages older than they are. I never had that feeling when I was teaching before, but I do now. I don't know how they show it, partly in the distance they keep and partly in other ways. I was bothered about this when I first went to that school, because I thought I was losing the intimate friendly relations I had had with students before.

This quotation rather neatly packages and illustrates a number of features of the student role in age-grading teachers. It is noteworthy that all teachers, married or unmarried, are aware of a change in their relationship to students. It is not merely a matter of students regarding the teacher as an "old maid"—it is more general than this. It is a matter of loss by the teacher of some kind of intimate, friendly contact with students, a matter of increased distance and formality. Students seem to age-grade teacher by age-distancing teachers away from them.

Teachers in their thirties were, among respondents, the most self-conscious and concerned about age-propelled changes in their relationships with students. At thirty-six years of age, 13s seemed to be reflecting intensely on this problem.

13s: I always start my youngsters out in the fall by saying they can call me anything they want as long as I don't hear it. I sometimes wonder how youngsters, how they actually picture me.

I do wonder about that, if I seem old to them, or just how they feel about my age. Every once in a while they ask me how old I am.

The change in student-teacher relations is not necessarily a one-way process. Another teacher in her thirties, 38s, expressed concern about becoming mechanical and "routinized" in her work.

38s: I look back when I first started teaching. I think maybe that was the happiest, because it was a new experience and I just went into it with everything. I think sometimes as we grow older we get careless and do things sort of mechanical. I think it makes your entire outlook different, and I would say maybe it happens around 35 years of age.

Teachers in the late forties and fifties also showed concern about aging but placed less emphasis on the loss of intimate, informal contacts with students and more emphasis on declining physical vigor.

> *4s:* I said to one of the other teachers just the other day—we both try very hard to push our shorthand students just as fast as they will be pushed and to do that we have to work awful hard—so I told her the other day I wondered if ten years from now we could do this. She said she wondered too.

On the whole, middle-aged teachers—in comparison with teachers in their thirties—seemed more secure and relaxed, less worried about their teaching, and considerably more subject to routinism. Apparently, in the course of becoming a parent-like authority figure, many teachers found that discipline problems decreased.

> *I.:* How do you think your life has changed in the last 10 years?
>
> *17s:* Well, I don't worry like I did. I'm more bored than I used to be. I don't worry as much. I used to come home and teach all night. One of the happiest things that has happened is that I can come home and forget my work. But I don't have the enthusiasm that I used to have, and I don't have the disciplining problems. I am about as peaceful now as I ever was.

There were some indications that this feeling of increased security and relaxation also included greater autonomy relative to the administration. While young teachers were dependent on and sensitive about principals because of desire for acceptance in the system, and most elderly teachers seem to cling to principals because of fear of change, some middle-aged teachers felt that they had become more indifferent and less vulnerable.

> *I.:* You said you thought you had gotten over some things. What do you think you have gotten over?
>
> *20s:* Well, I think I am less inclined to worry about other people's opinions of me than I was then. I remember when I began teaching, I had a fear of the principal and what he thought of me. Now I don't particularly care what he thinks. I would rather he liked me, but if he doesn't like me we can still carry on because it doesn't bother me a bit. Then, if I did something, I would wonder, "Now what is he going to think about that?" Now it is, "What am *I* going to think about it?" I think that I am very much less sensitive to other people than I used to be. Sometimes I am kind of sorry I do this in my work because sometimes it doesn't turn out. I mean I get a little indifferent and easy-going.

However, (as with teachers in their thirties) it seemed to be the students who reminded the teacher of her accumulating years. If, as in the case of 42s, a teacher began in the city system at the age of thirty, she might, in the early fifties, be teaching the children of her early students. This seemed to be a rather significant discovery.

42s: It's funny. I remember the earlier incidents in teaching better than I do more recent ones. I am now getting children of the students that I had when I first began to teach here. Every once in a while some child comes up to me and tells me that their father or mother said that I taught them. And they say that couldn't be because I am not that old, am I? Then I have to tell them that I am that old.

In general, there was among teachers in the late forties and fifties considerable concern about the long stretch ahead, the possibility of the burden becoming heavier, of interests declining, and of the distance between teacher and student increasing. Respondent 40s, fifty years of age, drew a distinction between being the age of the students' parents and being the age of their grandparents.

40s: I have a theory about this matter of teaching that I probably shouldn't state. I think that teachers retire too late in life. I think we should have voluntary retirement at fifty-five, put it on the basis of thirty years of experience if you wish, but I think fifty-five is the time to retire with compulsory at sixty. I feel that as long as I am the age of the students' parents, I should be their teacher, but when I am the age of their grandparents, I should step aside.

I.: Do you think that the student himself reacts differently to an older teacher?

40s: I think that a young teacher may lack in scholarship, but she has enthusiasm, and young people respond to youth. If a teacher is young and pretty, they like her, and it is easier for them to learn. . . . I think that our high school students assume that we are much older than we are. Being with young people does not make one young. There are times when being with them makes a greater gulf in ages really come true.

Similar views were expressed by a sixty-year-old teacher who advocated a retirement age at fifty-five although she was herself past that age.

35m: I don't think that I have the kind of influence that some teachers have. Part of it may be because I'm growing older. I think there is no doubt about it; it is difficult for a teacher to do effective teaching as she grows older. For the teacher's own sake, I think, heavy as the load is, the teacher should be allowed to retire at 55 if she wants to. Now they will let them retire at 62, if they have a health problem; at 65 if they want to. I'm past that 55 mark now. I think I'm going to have a health problem and retire at 62.

Most mature teachers did not explicitly spell out the possibility—so threatening to self-esteem—that they might not get the same response from students that younger teachers get. Yet, they seemed to be saying the same thing implicitly. When asked about the best teaching period, either for themselves or for teachers in general, they generally agreed that the early years were better, complained about the cumulative burden of teaching, and wished for earlier and better retirement arrangements.

I.: What do you think about most teachers—what is their best period?

23s: I don't know. I suppose maybe the first years that they teach, when they have more vitality. As you become older—I've heard so many teachers say this year, "Oh, if I just had enough money, I'd quit tomorrow." Teaching becomes, I've discovered, it becomes more difficult all the time. It may be a result of the war, I don't know. You don't get the type of work from the children that you used to. You don't find children that are interested in obtaining anything and it becomes a task—and I've heard that from many teachers. I believe it is when she has a lot of vitality and doesn't sink under the load that she has to bear.

The mature teacher's refrain—very general among those interviewed for this study—was one of frustration and helplessness, which alluded to administrative policies and the public attitude toward teachers, but which tended to be focused on students "getting worse every year." Teachers were not certain if this was due to their own aging or if it was involved with broad changes among students. In the manner of 23s, most teachers vacillated between two themes: "Teaching is for the young and vigorous" and "The younger generation is undisciplined, uninterested, morally decadent."

31w, a widow with grown children of her own, is an example of a teacher who concentrated on the second theme, placing blame on the times, the pace, and direction of civilization, on mothers for working and not looking after their children, and, to some extent, on the school system for lowering standards.

31w: The discipline problems are getting worse every year. That is true all over the country—not just here. I have a friend out in California who wrote me that she is so anxious for vacation this year because the discipline problems have been so bad.

I.: Why is this?

31w: Why? Because we are in an age of moral decadence. It is just part of the pattern everywhere in this country, in every line of work. It is a matter of getting by with what you can. It has permeated the whole country from top to bottom. The parents don't have the control over the children that they used to have. It is just our kind of civilization. The mothers aren't in the home much; they work or go out.

I remember when my children were young; they had a good time, but they didn't have to have a car of their own. They didn't do all the frantic rushing around and getting into trouble. . . . I don't know what anyone wants to be a teacher for. Not unless you feel the urge very strongly and wouldn't be happy doing anything else. I certainly wouldn't advise it for the money. And, I think, teaching is harder than it used to be.

I really wouldn't want to give advice to a young teacher. In my own field they have lowered the standards to such an extent. We just hit the high places. The whole trend has been to cater to those of lower abilities. We cheat young people who are superior. I often wonder if any of them are going into teaching. If they are, then are we giving them the right foundations?

I always try to have them read some Shakespeare. We used to give them "As You Like It," or some other light comedy when they were freshmen and then "Julius Caesar" when they were sophomores or juniors. Now, with my sophomores and juniors, I still have them read "Julius Caesar." Most of them, a lot of them, have never heard of Shakespeare before. It is just amazing how little they know. And their grammar when they are sophomores and juniors!

It should be noted that this statement from 31w contrasts markedly with her statement on experiences as a young teacher in a small-town school: "I had a wonderful time. . . . The pupils were just about as old as we were. We dated and had a wonderful time." It is reasonable to suppose that her bitterness about the current generation of children derives at least as much from lack of student interest in a woman of sixty-five as it does from lack of interest in the subject matter.

Types of Mature Teachers

The teacher quoted above is traditional in the sense that she is firm about discipline and, apparently, about presenting students with classical subject matter. This mode of adaptation appears to be a common one among mature teachers, not only because they were trained in an era of traditional education, but probably also because they are unable to function in group-worker fashion, and this is one of the few avenues open to them. It seems quite certain, however, that discipline remains a constant problem to 31w and that she remains generally dissatisfied with herself and the students.

However, the age distance between mature teachers and students does not eliminate the possibilities of workable relationships between them. There are mature teachers who continue to derive satisfaction from teaching and who seem to elicit interest and response from students—at least from a number of students.

> *44s:* I'll be straight with you. Most people think I'm the best English teacher in the city. My young people win many of the essay contests. I'm ambitious for them. If they dislike me it is because I push them so hard. I'm a pusher, an aggressive teacher. They don't encounter many teachers of this kind any more. Young people are not used to having people talk to them straight from the chin, but I feel that young people must have a feeling of success. I'm tough, not easy with them.
>
> *I.:* You said, if they dislike you . . .
>
> *44s:* I think if you drive them as hard as I do, some of them are bound to resent it; but I feel that it is worth it for the better students. A girl I had seven or eight years ago has been doing quite a bit of creative writing, has gotten quite a bit published. I think that is the pay-off. One of the nicest things is that years afterwards I will meet some of the students I have had on the street and they will tell me how much they appreciate my class, how they are going on with their reading. If I can stimulate that kind of interest I think I am doing a good job. . . . I do some writing myself. I have an article in the paper now and then, and I've written little things—travel sketches in *Scholastic Magazine* and things like that. I should do more writing than I do. I've never written for sale—perhaps I should. I'm not going anywhere this summer. I think I will stay at home and do a little writing.

The important difference between 31w and 44s is that the latter seems to lead with her subject matter, literature and writing.

She appears to keep discipline by making demands on students, as well as by being frank and brisk with them. For 31w, on the other hand, discipline seems to be an end in itself, and her central identity as teacher is that of a firm disciplinarian. 31w seems to conceive of herself as an English *teacher;* 44s conceives of herself as an *English* teacher.

Mature teachers who follow the 44s approach may be considered as a type characterized by a paramount interest in their subject matter and in teaching as a means of conveying subject matter and participating in a subject-matter field. Among the mature teachers in this sample there were only two who seemed to fit this model, 44s and 45s. The latter was a drama teacher who had been so active in producing plays and radio programs that she ran into difficulties with the administration. She had also maintained contact with drama circles and sponsored some students toward careers in the field.

Such teachers—ten of whom were known as "old characters"—seem to retain the respect of students partly because of extra-teacher status; they evince interest and engage in activities that establish them as a part of a professional community outside of the school system. In this way they appear to represent access to a fascinating part of the adult world.

Such teachers are likely also to have class status; they are likely to derive from upper- and upper-middle-class families and to have high status associations. 44s, for example, belongs to the Missouri Writer's Guild, the University Women's Club, and the organization of Business and Professional Women's Association.

There was considerable evidence in the interviews that this kind of mature teacher is held in high esteem by students.

I.: Do you know any outstanding older teachers that you feel are able to relate to the students?

40s: Yes, we had one retire from our school several years ago. It was a tremendous shock to us when she had to retire at seventy—we didn't guess her age by years. She was youthful in her appearance and in her attitude. The students liked her. She taught Latin, and she has gone into many activities with the D.A.R. It was a shame, from my point of view, for her to retire, but frankly, I have known a good many teachers who have not reached the age of retirement who should have the privilege.

I.: Now, what was she like?

40s: In the first place she was pretty. She had white curly short hair, and bright eyes, and a lovely complexion, and tremendous enthusiasm. She told me once, that when she was a young teacher and woke and found that it was Saturday, she was disappointed because she couldn't teach. She is one of the few teachers that I have ever known in my life who were disappointed when they couldn't go to school. She had that attitude, she entered into the lives of her students, she had a good time with them.

Mature teachers of this kind apparently were more abundant in the previous generation. Most of the comments were about retired teachers. From these comments it appears that, as teachers in their thirties and early forties face the problem of age-propelled changes in their relationships with students, they sometimes look to such older retired teachers as possible models.

I.: How do you think your generation of teachers compares with the older generation of teachers?

38s: Well, I have a great deal of respect for the older teachers because I can remember some of the teachers that I had. I taught on the same faculty with some of them. I have a great deal of respect for their thinking and learning. I think with some of the modern ideas—I feel that sometimes we are too lax. I feel that quite keenly at times. Oh, I don't believe in being too hard with a child, nor too lax. I still say that you can be firm, show a firm ground and be fair. I know that with my experience with the students that come in and say, "Oh, you can get by with murder with so-and-so." They have no respect for that type—I mean that type of getting by. I have heard students say, "Oh, I would never trade all of my years for the time that I had in Miss ———'s class," or "I certainly did learn a lot from Miss ———." Miss ——— is one of our retired teachers. Oh, a keen mind; I wish I had her vocabulary or just one-tenth of it. Even to this day she's very alert. Now, in her last year of teaching, she was not hearing everything, she wore a hearing aid—maybe some of the youngsters thought they were getting by with things in her class, but she knew what was going on. You'd be surprised at the number of students who would say, "If you want to learn, you'll learn in Miss ———'s class, because she's teaching it." There again, it's that age that makes you feel that they're firm and yet again fair.

It appears that mature teachers can achieve firm discipline, provided that discipline is introduced for the sake of teaching subject matter and not for the sake of punishing or controlling. Even "old character" attributes are tolerated by students and by young colleagues if the mature teacher is known as a good teacher in the matter of presenting subject matter.

13s: One of our favorite teachers—no one knows how old she is, but she dresses queer, yarn hats and so on. But she is an outstanding teacher and those youngsters just think she is the finest. But to look at her you wouldn't know that she is so well-liked by the youngsters. She has a marvelous sense of humor too. I think she has been teaching since I was born.

There is, in addition to the "old character" style, another way in which mature teachers can bridge the age distance between themselves and their students. Some mature teachers relate to students by making use of their motherly or grandmotherly qualities, by being generally kind and understanding, by functioning as counselor as well as teacher.

19s, a commerce teacher, fifty-eight years of age, is an example of the mother-counselor type. It is interesting to note that the mother-substitute function is consciously adopted and cultivated.

19s: Some of the things that we go into in this particular class that I have—we try to take up some personality training and talk about how to improve our appearance. We discuss social problems, and even how to act on dates, just everything that comes up. We don't stick to any textbook at all. Problems that come up with a group of girls—you just work it out any way you can. We preface our remarks by saying, "This might happen to you when you're on the job and would be very embarrassing. Would you rather iron it out now or have the boss iron it for you?" Then we go into the situation, something someone has done that might be embarrassing on the job. That's why I say they sort of look on me as a mother. They come to me with some of the funniest problems you ever heard of. There is no limit to what we can talk about, you know, just a bunch of girls.

Another outstanding example of a mother-counselor type of mature teacher is 56m, a sixty-year-old home economics teacher, who had convinced the administration that a family life course should be set up for her to teach.

56m: When the class first begins, I tell the girls, "This is the place to ask questions you can't ask elsewhere." I encourage them to speak up. I thought they were asking about all the questions they had on their minds, and then one day one of them asked that we get up a question box. And then I just got a pouring out of questions.

> Mostly I just serve as a sounding board for them, and kind of a counselor. Actually I don't do very much teaching in the usual sense. We have little tests and things of that kind. They get kind of noisy sometimes, not unruly; I let them go as long as they do it because they are interested in what we are discussing.

The mother-counselor variety of old teacher seems to be most adaptable to such "nonsolid" vocational courses as commerce and home economics. It may be the result of an application of progressive methods by mature teachers to classes of girls. 56m implies that it may not be adaptable to mixed classes or to solid-subject classes.

> *56m:* I would never advise a young person to go into teaching. They might not land in as good a place as I have now. Some totally undisciplined children are being sent to us now. They get worse every year. I don't know why exactly; it is the way the parents bring them up, or don't bring them up. And a lot of outside influences. They have no standards of behavior. I'm pretty lucky in dealing with the older girls. There is more trouble in the academic courses.

The evidence is considerable that, beginning in the thirties, there is a decline in intimacy with students. Some mature teachers (by virtue of a keen sustained interest in their subject matter, accompanied by an established prestige) seem to be able to maintain effective relationships with students. A few, in special situations, develop intimacy through a modified progressive approach.

In the opinion of a large number of mature teachers, the students "get worse every year," the teaching load becomes more burdensome, and the prospect of retirement becomes more attractive.

Age and Generational Differences in Educational Ideology

Teachers of such vocational subjects as commerce and home economics seem to have adopted progressive methods more readily than teachers of such "solids" as English and mathematics. Teachers who derive from upper-middle-class and upper-class families—and from cities—seem a little more inclined to teach the traditional academic subjects and to adhere to traditional methods, than those who derive from small towns and rural areas and from lower-middle-class and lower-class families. The latter are somewhat more likely to be found teaching the vocational subjects and are more likely to follow what they consider to be the progressive approach. No doubt personality patterns also are involved.

However, age-generational differences is by far the most important determinant. Such mature mother-counselor teachers as 19s and 57m are *avant-garde;* their age peers are overwhelmingly traditional. Although long exposed to progressive education in summer school and through educational literature, most mature teachers are defensively traditional, either because as young teachers they were set in such a mold or because as their age distance from students increased, they abandoned what progressive experimentation they had attempted.

The young teachers in the sample—those in the thirties and early forties—were, of course, trained during a period in which progressive education was in vogue. These were concerned about the age-associated loss of intimate group-leader contacts with students, worried about the feasibility of progressive education in large schools (characterized by large classes of short duration), and used teaching methods that probably would not be considered progressive by exponents of the movement (especially those who are now employing group dynamics). They did, however, align themselves with progressive ideology.

One aspect of this situation is that the two ideologies seem to serve as rallying points and are used for name-calling purposes in the conflict between young and old teachers. This conflict, typically, concerns the promotion and defense of seniority privileges by mature teachers and young teachers' efforts to secure increased status for themselves. The fact that methods used by other teachers are cause for concern (despite the formal administrative separation of teachers as terminal staff workers and the physical separation of classrooms) indicates the degree of a teacher's identification with the occupational group. The fact that teaching methods serve as platforms in intergenerational conflict indicates the depth of the teachers' involvement with students.

I.: How would you say that your generation of teachers compares with the older generation? Are they different?

2s: Um-hum. Yes, I think so. I think there's quite a difference. We are more interested in children as children, they are more interested in getting the subject matter across. The older teachers want the children to *sit* there, and *behave* themselves, and *learn* their lesson, or *else!* The younger teachers are more interested in children as children. Oh, don't get the idea that we aren't interested in the subject matter. We want them to get so interested they will have a thirst for it. And don't get the idea that the older teachers are all like that. I've had some myself, this journalism teacher I was telling you about, who got us interested in poetry. She was a very good teacher.

Most young teachers criticize their older colleagues for being rigid disciplinarians and having lost (or never having had) an understanding of the students. Like 2s's statement, "Oh, don't get the idea that we aren't interested in the subject matter," there is usually a defense against the charge of holding loose academic standards. However, perhaps because of the power that mature teachers hold in the schools or perhaps because of cultural rules calling for respect for the aged, young teachers usually qualify their objections to the older generation with, "Don't get the idea that the older teachers are all like that."

The young teachers' attitudes toward older teachers reveal a factor that stands apart from concern for students and that involves them as members of the occupation. Mature teachers sometimes are seen as representing what they fear the public thinks of all teachers and what they fear they may become as they grow older.

I.: Have you noticed much of a difference between the older and younger teachers?

12s: Let me approach it this way. I think . . . many of them have become—well how to put it—old-fashioned, which twenty years from now someone will be saying about me, I guess.

I.: What do you mean old-fashioned?

12s: A lot of them seem to forget that they themselves did many of the same things when they were young, the same things that a lot of these kids do in high school. They don't do mean things, at least not many do, but just little things. The older teachers get so they can't take it. They get on her nerves and then she is tromping on someone, and that isn't the way to handle the kid at all. Many of them have the idea —I don't mean they have the idea—they have been, have had to be in control of their classroom for so long, and it sort of carries over into other things, subconsciously. And that is why a schoolteacher nine times out of ten looks and acts like a schoolteacher.

The most important elements in the young teachers' battery of criticisms are included in the statements of 12s and 2s. According to them old teachers are not interested in children as children; they are so far away from their own youth that they have forgotten what it was like to be young; they confirm the adverse public stereotype of teacher; and they do not carry their share of the extracurricular work load.

Among the teachers interviewed for this study, those thirty to thirty-five years of age were the most critical of mature teachers. Although teachers younger than thirty were not interviewed, it can be assumed that they fit into the same pattern (especially in view of the mature teachers' attitudes toward them).

Teachers in the late thirties or early forties took a more moderate stand. One, 14s, thirty-eight years of age, identified herself with older teachers. As she describes it, this occurred after she had attempted to be progressive and had found progressive methods not feasible in the structural arrangement of the large public high school

I.: Do you think there is a special difference between your generation of teachers and the older generation?

14s: I think I am of the older generation of teachers. That is, I think my ideas belong to the older generation. The ones I've had contact with, I think I find myself pretty much in agreement; you know, the traditional old-fashioned way of teaching. I believe in discipline. Oh, I probably give more freedom in my speech classes, but those are the kind of classes you give freedom in. But I don't think I'm much different from the older teachers.

I guess there are schools where they are more progressive than we are here. Actually, in high school, the high schools are not set up to be progressive. When you have students running from one class to another, you fall back into the old ways of doing things. You have to check them and get a grade for them. It is very traditional. I found that out in this drama class. I had to give grades and there wasn't any basis for giving grades except for my own judgment so I had to make sure every student had a fair chance to perform and that there were a lot of students, so it had to be organized more.

It seems significant that the "generational difference" question was invariably interpreted as calling for a stand on differences in teaching methods and for an evaluation of them. In a sense, this question also caused the teacher to age grade herself.

14s's identification with the older generation was associated with an effort to withdraw from drama teaching. There was noted also a "getting nowhere" kind of demoralization.

I.: What do you think your life will be like ten years from now?

14s: Gee, I don't know. You shouldn't ask that. I don't think it will be much different, except that I will be older. In fact, I can't see that it has ever been much different. Of course, I'm not even one who thinks it would be bad to retire. In my present state of mind I think it would be all right. Just sit around and do nothing.

I.: What was the happiest period of your life?

14s: I think probably when I taught in the elementary school upstate. I think that was probably the happiest. I enjoyed the work and—I guess it wasn't the work, I guess I had a better time. There were a few young people and we could go places.

14s seemed to be engaged in making a crucial age-identification transition from a young woman in the marriage market to a resigned old-spinster teacher. It appears that identification with the older generations' teaching methods serves as a medium for changing identification.

It has been pointed out that a basic change in relating to students seems to occur as teachers approach middle age. From the standpoint of the teacher's more comprehensive status as a woman, the nearly final withdrawal (or exclusion) from the marriage market may represent a more fundamental change. Where chronological age was specified, resignation to spinsterhood—like resignation to status of parentlike authority figures—was placed at thirty-five years of age.

I.: If you had your life to live over, what would you do differently?

42s: Oh, I think I would get married if I had a free choice. Not just anybody, of course, but I think marriage is the ideal for women.

I.: How long have you felt that way?

42s: Oh, I guess from about thirty-five on. Up to that time I just sort of assumed I would get married. Somehow it just didn't work out that way.

Resignation to spinsterhood seems to be a gradual, cumulative process, closely involved with increasing commitments to an advancing career line. It also appears to be involved with movement of teachers from rural areas and small towns to city systems. There were, moreover, some teachers who seemed to have never seriously considered marriage. But, for all, there seems to be a period, at about thirty-five, when there accurs a nearly complete acceptance of a mold that has been set in the past.

42s: When you have given up hope of ever getting married, not until then are you completely satisfied with the job of being a teacher. I will be quite frank with you, I would much rather have married than to teach school, but I still wouldn't like to combine the two jobs either. It is very hard for me to say when a professional woman is settled on the job. I suppose it is when she is about thirty-five, and isn't so interested in getting married anymore.

Whatever the age of resignation to spinsterhood, the middle-aged teachers in the sample seemed the most settled and satisfied, the most committed and, according to their own criteria, the most professional.

I.: How do you think your generation of teachers compared with the older generation?

3s: I think there are more of us who are teaching as a way of making a living. I think a lot of the older teachers taught just because it was open to them; a lot of them had the attitude that it was just something to do. I mean, they didn't have to teach and they did have a kind of missionary spirit, but I think my generation is in a way more professional. It's a way of making a living and there are probably more of us who come from the lower- and middle-income groups.

I.: What is a professional teacher?

3s: To me one of the requirements of a professional teacher is that it *is* your job and that's what you're prepared to do and that's what you do and it's not a steppingstone to something else, not a place to put in your time until you get married or something better comes along.

3s can be considered a representative spokesman for the middle-aged group. Generally middle-of-the-road on the progressive-traditional issue, they employ (in somewhat moderated form) both the young teachers' criticisms against mature teachers and mature teachers' criticisms against young teachers. They are somewhat envious of the older generation's higher class origins and "kind of excellence" in teaching. They are equally envious of the greater economic, educational, and job-placement opportunities available to the younger generation.

4s: The on-coming teachers have a better opportunity to get into a much better position than you thought you could ever have done here. They wouldn't even consider hiring you in the city system unless you had at least three years of experience. Now we take them in high school right out of college. We don't want to but that is the way we have to do. And then most of them can go right on through college and most of us didn't have that kind of money—times were harder then—and we had to go to school and teach. I don't know whether they appreciate the opportunities they have; most of them seem to be very casual about it.

Middle-aged teachers generally viewed themselves as the strong, stable backbone of the teaching staff. This is partially generational in the sense that they view themselves as the "hard knock" generation, the group that fought a difficult mobility route in an era when teachers were in surplus and requirements were raised higher and higher. To some degree, it is also age graded in the pure sense: middle-aged teachers feel they bear the heaviest load because young teachers are inexperienced, and more important, uncommitted; old teachers tend to withdraw from auxiliary responsibilities.

5s: Well, I think we're more serious about our work, maybe more concerned about doing a good job and we've been doing it so long that we know what it is that we're going to finish. So many of the young people stay in only two or three years before they get married, and I don't think they are quite as interested in the work outside the classroom as the teachers who have been doing it a good while and plan to finish out their career in teaching. The middle-aged group, I believe, is the one that has the biggest burden of the load. The older ones are given fewer and fewer extra responsibilities and the younger ones necessarily, because of lack of experience, are given a much lighter load—that's why I believe the main burden rests with the middle-aged group.

Age and Generational Differences in Extracurricular Activities

Middle-aged persons seem to carry the largest proportion of the load in most work organizations and voluntary organizations in our society. In schools, the central function of teaching is divided among all teachers, but the auxiliary functions, such as extracurricular school activities and the affairs of professional associations, seem to fall heavily on the middle-aged. This can be viewed as resulting from processes occurring on two levels—the organizational setting and the personal career. The middle-aged teacher is thought to be most available. In the schools, as in most organizations, this is a result of many acquaintances and greater familiarity with program. With time, it comes to be assumed that the teacher is satisfactorily performing the central teaching function, that she is generally reliable, that she can afford the time for and can be trusted with other functions.

In addition to organizational factors, critical early changes in the teacher's career seem to have the effect of making middle-aged teachers susceptible and responsive to auxiliary responsibilities. As contacts with students become more formal and less intimate, and as spinsterhood comes to be accepted, the teacher becomes more deeply committed to the teaching career, more professional, more thoroughly colleague oriented. In interteacher relationships, this tends to separate members and creates a strong tendency to reject young teachers as not committed, not legitimate full members.

This lack of acceptance is seen in several areas, among them the professional association. Teachers' organizations in the city studied can be divided into two types. There are (1) professional associations organized within the "cooperative council" or affiliated with the state education association and the NEA; and (2) two unionlike interest-group organizations from which administrators are excluded. The NEA affiliate (often called the "company union" by teacher members) offers membership to any teacher in the school system. However, judging from the age of sample teachers who are now active, and from comments about others who are active, it appears that members who hold the most responsibilities are middle-aged, and those who hold offices tend toward being late middle-aged.

In the case of the two interest-group organizations, the middle-aged also carry the main burden of responsibility. These, however, make an important addition to the age-grading pattern. Although neither organization has succeeded in recruiting more than a minority of teachers, and although they compete with one another for membership and with the NEA affiliate for loyalties of members, they do not consider young teachers to be desirable additions.

4s: A lot of young people that come into the system have no idea of feelings. They don't care whether you have retirement or whether you have a pension, or what is going to happen ten or fifteen years from now; they are just interested in what's right now and that's all. For that individual, the union, you see, is of no value to them. With your young people you get a lot of them with that feeling.

In the A.F.T., we prefer to hold back a bit to see what they are going to be and then ask them. We invite them to join and then they are voted in. In the other organizations, all they have to do is just pay a fee and then go in. While we want new members, we prefer to see what they are going to be like before we take them in.

I.: You said that you wanted to wait and see what kind of people they were. What kind are you looking for?

4s: We are interested in two things in particular, I would say, and one is that we want someone that is interested in teaching and in making teaching a profession and the teaching career a better career and is also interested in having a better school and better situation for the students, instead of just for themselves.

Respondent 4s, forty-eight years of age, is active in a number of commerce-teacher activities as well as being a leader in the American Federation of Teachers. While such functions may be viewed as auxiliary to classroom teaching, they seem to be vital to the collective solidarity of the colleague group and perhaps also vital in providing career-committed middle-aged teachers with leadership opportunities to compensate, in part, for other career ceilings. The fact that the younger generation of teachers is not turning out to be led seems to be responsible for some of the resentment of middle-aged teachers toward them.

16s: They're not dedicated to their work at all. They're not going to do anything extra—activities or professional organizations or anything else. When the bell rings, they check out—school's out. Now, we were brought up to feel that we were getting a lot out of it because we were putting a lot into it. I don't think they feel that way at all. It's just a job to them. I feel that they don't have the same devotion to the teaching profession that we have.

Like the middle-aged, older teachers are concerned about young teachers' lack of commitment and dedication, but they also are greatly concerned about the younger generation's preparation for teaching and methods of teaching. Many older teachers, because they hold firmly to traditional education, feel bitter and abandoned. Young teachers, along with students and society at large, have moved away from the cherished, and now somewhat sanctified, values of their youth.

I.: How do you think your generation of teachers compares with the younger generation of teachers?

48s: There is a great deal of difference. . . . I hate to sit in judgment on others, but I don't think they have been held to the standards themselves, these teachers who have been coming out of the schools now, not the standards that we were held to. I have teachers coming from the teachers' college to do practice teaching. I can see that they haven't been held to do things just so, or maybe they are products of the war too. They aren't good spellers; they can't go to the board and write out a sentence for the youngsters and spell correctly. And some of the colleges are saying, "What if they can't spell, some of the smartest people can't spell." Well, I say that if they can't, they had better learn, and things like that, you know—being held to correctness. I just thank teachers who did so much to wake me up. I don't know whether we will ever get back to that standard or not. I don't know.

All of this reflects the identification that old teachers have made, through the years, with the field of education and with the teaching profession. These are people who have long been committed and who, by virtue of this and of energy expended, have a large investment in the profession. In the course of time, sacred qualities are attached to the conditions and procedures of the school, the forms of organization and interaction in the work setting, and the model of teacher and teacher career.

In the work system that stresses developmental training, some characteristics and motivations of young teachers are placed, by mature teachers, in a special career context. Young teachers are sometimes viewed as inadequate because they do not follow the traditional rural route and because they are thought to be inadequately trained by the colleges.

I.: How do you think your generation of teachers compares with the younger generation?

8s: Do you want me to be frank? I think that we are much better prepared. I can prove it this way. Too many of our younger teachers can't teach but one thing, and I don't think that is wise.

I.: You don't think they should specialize?

8s: Absolutely not. I would advise them to start in a rural school or a small town . . . that is good experience, but not many would want to do it. Now they want to specialize, to teach one subject, and that is why a lot of them are not a success—their education is not broad enough. I shall never cease to be grateful to our college president who insisted on all of us being broad in our education. When I was teaching in small schools I had to teach math, art, music, and many other things. I think in the old days we had more pride if we could do more than one thing.

For the old unmarried teacher, the recent acceptance of marriage as a part of the teacher's career line is the most radical change of all. Perhaps, because of an awareness (threatening to self-esteem) that devotion to the teaching profession has had the effect of closing out full womanhood experiences, the young teacher's interest in marriage receives special attention. In the interviewing, it appeared that the greater the affective pitch of the criticism of young teachers, the more likely it was to include reference to the young unmarried teacher's interests in dating or to the young married teacher's interests in husband and family.

9s: Our experience in the past has been unmarried girls come into the profession, and now we have the married girls coming in. It has just been since World War II. Before that, you forfeited your position. I think that it isn't probably a question of preparation or ability or anything that we feel differently, but you know yourself that your interest is divided if you have a home, a husband, and teach. If you aren't married, you are perfectly free to go to all the meetings and attend them and join the different school organizations. There is that difference in the girls with the home and husband to think about and her having to curtail things she does outside of the schoolroom because of that and the teacher who is free to do and go and take part. The married girl is not so apt to take part and belong to the organizations, those professional things that go along with teaching.

The advent of young married teachers threatens to reverse the process of professionalism that represents the teacher's identification with the occupational group. The acceptance of married women as teachers has the effect of diminishing the worth of the old unmarried teacher's more complete commitment and of calling into question the wisdom of her earlier decisions.

The interview with 9s is unusually revealing. Having voiced such extreme objections to married teachers, later in the interview she gives an unusually blunt statement of her own regrets about not having married.

I.: If you had your life to live over, is there anything you would have done differently?

9s: I probably would have gotten married. If my experience now would count then, I probably would have done that differently, but as far as the profession, I still would have chosen the school.

I.: Was there any particular reason why you didn't get married?

9s: It was because of the responsibility for my mother. At that time you couldn't ask a man just starting out to take on a family of two. And in the school here you couldn't teach and get married too. Since that rule changed, it has made it lots easier for lots of teachers. You would be surprised how many of the teachers have someone to take care of, especially the women. Men don't take to responsibility.

Speaking from an advanced position in life (sixty-seven), 9s still places primary emphasis on her occupational identity. She would like to have taught and married—"as far as the profession, I still would have chosen the school." Among her life commitments, teaching seems to have come first, her mother a close second, and the model of marriage third. That she should give primary emphasis to teaching seems especially significant since, among the mature unmarried teachers, she was one of the frankest in her regrets about not marrying. The majority conveyed a sense of stronger commitment to teaching than to other interests established in early career.

Among the teachers in the sample, one had married for the first time at the age of fifty-five. Even in this case (although she did marry and although she was excited about and apparently gained prestige and self-esteem through marriage), the description of preparing for marriage suggests that teaching remained primary in her life.

35m: You see, I've only been married five years. I wasn't young when I got married. Before I got married, I consulted Mr. ——— he's in charge of personnel downtown—and I consulted my minister and doctor. Mr. ——— said, "There'll be a position for you as long as you want to teach." He said that the trouble with so many older teachers is that they have no outside interests. He said that it was nice to know that I would have outside interests and that I would be a better teacher for it.

The fact that marriage should be defined as an outside interest, by both teacher and personnel director, may be viewed as symbolic of the marriagelike quality of the mature teacher's commitment to the occupation—a commitment that sets her apart from young teachers and that sets the tone of interteacher relationships along an age-generation axis.

Sponsorship, Intergenerational Conflict, and Age-Graded Cliques

In portions of the interviews dealing with career experiences, old teachers loomed important, and often forbidding, in young teachers' experiences in the city system. Mature colleagues seem to have been the most influential of all agents in the incoming teacher's world. In some cases this was in the form of guidance and sponsorship, but in most cases it involved ignoring, hazing, or blocking the newcomer.

These forms of interage relationships among teachers are not confined to the young teacher's first year. In a system staffed with large numbers of mature single women, teachers may be considered young for some time—until they reach the late thirties or early forties. It appears that a young teacher is most likely to continue to be viewed as young and in need of help if she accepts initial guidance and sponsorship from mature teachers.

I.: How did you feel the first year you were there?

37s: I wouldn't be a very good example of that in general reactions because one of the teachers out there had known me since I was six. . . . Of course, she gave me a wonderful send-off with the principal, and they all sort of hovered around me as if I were about ready to fall apart. They didn't think I was quite bright I guess. And they still act that way. It is kind of hard for them to realize that I am grown up and quite capable, so I just listen to them and then go my own way.

Respondent 37s is now thirty-six years of age and has been teaching at the school for six years. When she entered the school, at the age of 30, she would not have been considered young in most adult circles. Her view of herself as not a "good example" seems to imply that being ignored by mature teachers or having conflicts with them is normal in the high school. At the conclusion of the passage she intimates that if she did not accept this relationship, she might be faced with alienation. Elsewhere in the interview, she indicates that she is the youngest teacher in the school.

It happens that one of 37s's motherly colleagues is 35m, an older teacher who was also interviewed in the study.

> *35m:* When she [37s] first came here—we are in the same department—she wanted to know about things, and I told her I would be glad to help her. She said she was concerned about taking too much of my time, but I told her that I was just passing on what was passed on to me. She wanted the children to have confidence in her, and she thought it would help if they were on the same pages I was.

In her approach to young colleagues, as in her approach to students, 35m functions as a mother counselor. In this, she is drawing from her own experience—her career being the most sponsor-laden among those of teachers interviewed. A high school principal influenced her to enter teaching, a small-town banker financed her education, a state teachers college superintendent intervened to get her a job and, in the city system, she developed a close mother-daughter relationship with an older teacher.

> *35m:* She was a teacher when I was there and became principal after I left—has remained a very dear friend of mine all these years. She died this year, died in April. She was eighty years old. That was just after they had had a surprise party on her eightieth birthday. You may have read about it in the paper.
>
> She inspired so many people! I had good training there under her when I first came here to teach. She was teaching the seventh grade then.
>
> *I.:* How did she help you?
>
> *35m:* Well, that is pretty hard to say. She was just very fine and sincere, well-liked. She was a very good friend of mine. I knew she was a very good friend of mine.
>
> She was quiet, unassuming, real, scholarly, just about everything that a teacher should be. Four years after I came here my mother died. She stepped right in then. She was a lot like my mother. She served as auntie and mother to me.

This kind of guidance sponsorship has strong and overt kinship qualities. When 37s was placed in a school with teacher friends of her mother's (shortly after her mother's death), her colleagues took it upon themselves to serve as mother to her. In teacher colleague relations, one of the strong tendencies is that of establishing kinshiplike cross-generational ties. In families that include female teachers of two or more generations, occupational colleagueship reinforces kinship to produce a family of adults with, for our society, unusually strong ties. For example, 18s (in whose family "teaching is quite an old profession"), now lives with a sister who is a grade school principal, an older widowed sister who was once a teacher, and an aged mother who taught school in her youth. For this family of aging adults, practically all outside contacts are limited to the teacher world.

Perhaps because of a need on the part of old people to maintain contact with the growing-developing younger generation, unmarried old teachers seek to extend their services as mothers or aunts to young teachers. Young teachers seem to reject such offers unless unusually immature or lonely. The cumulative wounds of such rejections may result in rigid antagonism among old teachers and a collective intergenerational hostility.

In any case, among teachers as a whole, intragenerational ties seem much stronger than intergenerational ties. In most descriptions of early experiences in the city system, there were accounts of uncomfortable situations with mature teachers. These extended all the way from concern about being ignored, or "cold shouldered," to open conflict. It seems to be partly because of such difficulties that young teachers form into generational cliques.

Most of the cases of conflict reported in the interviews involved an older and a young teacher in a situation carrying implications that, in standing up for her rights, the young teacher was fighting a cause for her set—while the old teacher was a leader of "old guardism" in the school.

42s: Well, the school was very crowded the first year I was there. Since I was new I had to change rooms and sort of fit in where no one else was using space. There was one older teacher who was quite a character. She was the one who gave us the most trouble. There were about eight of us who were the same age as I was. She used to call us "the brats." Let's see, I was twenty-three then. But we were young enough so that the janitor would stop us to prevent us from going places where the students weren't supposed to go. Anyway, I never really got along there until one time I stood up to this older teacher and after that she had respect for me.

One of the rooms I had was next to the physics lab and it just wasn't large enough for my class. Finally I found out that the reason I had the small inconvenient room was because the older teacher was using one down the hall, and had insisted on having it. And that it was purely for her personal convenience; in fact, the room was vacant sometimes and I wasn't assigned to it.

So finally I went to the vice-principal and he supported the older teacher. He said that this was being done out of respect for the older teacher. I told him that I thought things might be arranged in such a way as to suit the needs of the students for space.

So finally this had gone so far and so many people were buzzing around about it and everything that I went to the superintendent and he backed me up. I had risked offending two persons—the vice-principal and the older teacher. But after that the older teachers were quite nice to me and I never had any more trouble. Apparently it was necessary to do something like that in order to gain their respect.

A mathematics teacher, 42s, now fifty-three years of age, is a spunky, individualistic person who led one of the union-type teachers associations in a successful drive for salary increases. It appears that such a person, while still young, may gain a position of relative equality with mature teachers. Young teachers who accept motherly guidance are treated as young and immature for some time.

Although 42s indicates that she gained the respect of older teachers in the school by standing up for her rights, it is not clear whether or not she helped the position of other young teachers. The vice-principal's stand is indicative of the importance placed on seniority rights for the "old guard" in the administration of the high school. Yet it appears that when a mature teacher carries seniority rights too far she may lose face and support. In a similar conflict over room assignments, when the old teacher had lost the initial case, other teachers banded together in an effort to strip her of other privileges.

23s: I had been having trouble with my eyes. The oculist I was going to was the same one who had the principal under his care, and when I had been there for a while my eyes were getting very bad. And, since I was new, I had been put into a room that had been part of an art room, and they had cut it up to make more classrooms; so this was a little room that had only light from the skylight and very tiny windows up at the top of the room. My eyes were bothering me terribly, and this oculist asked me what was the matter, so I told him. So he called the principal and told him I would have to be put in another room.

And then the principal put me in a room an older teacher had had. And she just blew her top. She was tremendously angry that a new teacher had been put in her room. She had moved out, she didn't want it, but the idea that somebody new in the system had been given that room was just more than she could bear. She fussed to me, she fussed to the principal.

Two or three other teachers came to me after that and wanted me to complain that when this teacher had moved out of the room they gave me she had more bookcases than she should have. Well, I didn't do it. I had no particular axe to grind about how many bookcases she had—it made no difference to me.

It is quite evident that—in an occupation characterized by equality of office and an early career ceiling—such seemingly minor privileges as room assignments have become important prestige symbols for old teachers. Although the assignment of rooms on the basis of seniority may be overruled by other considerations, and although some old teachers may overplay their rights, old teachers seem to maintain an age-graded or generational front on such issues. Faced with a common state of "disprivilege" (as well as other common interests) young teachers seem to form cliques and hold to them for many years.

The most serious conflict was reported by a young married journalism teacher who, apparently without being aware of it, committed an act that old teachers in the school interpreted as evidence of great disrespect for the life investment of mature teachers.

52m: But something did happen this year. This year a teacher who had retired died. She had been quite friendly to me. I didn't know her well; I didn't know any of them well. But we put a brief obituary in the paper about her death. We might have done more with it but the students who were writing it didn't know her. It was a regular obituary item.

Well, at lunch one of the older teachers, the so-called head of the

English department, landed on me. She was just simply furious about this. She was very overwrought about the whole thing.

And the thing that seemed to bother her most was that we had printed her age. This teacher that died was seventy-one, respectable age to die, really. I didn't know whether they thought that her age should not have been disclosed because she was too young to die or too old to have just been retired or what. The kids had gotten her age from an obituary in the local newspaper. . . . Well that started things off. There was an undercurrent all through the school. I'm sure that each of these older teachers read this as her own obituary. They thought, "When I am dead and gone, after serving all these years, is this all the recognition I will get?" . . .

There is a group of older teachers out there that we call the twenty-seven year club. They are always talking about their twenty-seven years of service to the high school. Most of them, of course, lined up with this older teacher who had started it. Some of them mentioned their indignation in their classes. I know that one in particular did it.

And the principal is halfway scared of her. Eventually he showed up with an editorial she had written in praise of this teacher that had died. He asked me to print it. I just told him that this had gone too far, and all over nothing; that I would quit before I printed it. And I really would have. I would have been glad to have printed it if she had come around quietly after seeing the obituary and said that she wanted to write an editorial for the next issue. . . .

Well, this thing over the obituary just dragged on and on. My lunch hour was the same as this older teacher that caused all the trouble, or started it. I would go down to lunch and grit my teeth and eat at the same table with her. Neither of us were going to back down on that. But it was kind of nerve-racking. And not very good on the digestion. We never spoke to each other but we always ate at the same table. It was all so silly. In the first place, we print the paper more for the kids than for the faculty. And the kids didn't demand a long, flowery editorial.

This story is an extreme example of intergenerational conflict. It is an incident that involves a threat to the service and sacrifice-based recognition claims of old teachers and one that demonstrates their manner of reacting toward those who threaten such claims. It derives from a convergence of conditions. The "twenty-seven-year club" probably has an importance in the school because its members have taught there since the building was dedicated. Membership probably represents a special sense of sacrifice because it is a working-class school. As she observes, 52m is unusually vulnerable because she is both married *and* young, because the school paper is exposed for all to see (a point she did not bring out), and because—judging from other parts of the interview—she inspires the interest of students as few mature teachers do.

Most interteacher conflicts reported in the interviews were between "young" and "old" teachers. The fact that middle-aged teachers generally were not reported as involved in interteacher conflicts may be due to some inadequacies in the interviewing. The interviewer probed at some length into experiences at the time of entering the city system and, as one teacher mentioned, "It is funny; you know, I remember the earlier incidents much better than the later ones."

On the other hand, it may be that teachers during middle-age are somewhat removed from conflicts of this kind. It may be that the middle-aged take the "middle-of-the-road" in disputes between old and young teachers—as with their attitudes toward educational ideology.

That conflicts should take an old-versus-young form is associated with several conditions pertaining to teachers: intergenerational differences in educational ideology and in mode of commitment to the occupation; age differences involving response from students; and, in this particular city system, an unusually high proportion of mature female teachers in the high school. In addition, the formal equality of teacher offices, the staff-career setting, and the early career ceiling may be conducive to this kind of conflict. Under such conditions, old teachers may be inclined to place emphasis on informal seniority rights that often impede the functioning of the organization.

Although the staff-type character of teacher offices have been stressed, there are some minor line-type offices in the form of departmental chairmanships. These offices usually have no special authority but involve some recording, reporting, and arranging of departmental affairs. Judging from the interviews, teachers do not feel that departmental chairmanships carry any particular prestige. Also, they involve additional paper work, which most teachers resent and despise. Chairmanships seem slightly more important in such departments as physical education, commerce, and home economics, which require special equipment and arrangements and in which the teachers work more closely with one another.

The one instance of line-type conflict reported in the inverviews involved a sixty-two-year-old departmental chairman, a rather vigorous forty-four-year-old teacher, and a thirty-year-old married teacher. Although both were interviewed, only the young married teacher reported the situation.

> *33m:* It has been a difficult situation there. I didn't know what kind of a situation I was getting into. I've never seen anything like it. Well, there is one teacher who is about forty and who has been there quite a while and then there is another one who is older and she has been there a long, long time. I was told when I went there that a teacher never stayed for more than a year. I've been there three years but it hasn't been easy. I guess that's about all I should say.
>
> *I.:* Do you mean they are in competition?
>
> *33m:* Yes, that's it. They're both anxious to win the favor of the students and to run the show. I've never heard of anything like that being so much out in the open. And I'm caught in the middle. There isn't much I can do except try to be neutral in a wishy-washy kind of way. And that isn't easy for me to try to do.
>
> When I first went out there, the supervisor downtown told me there isn't a departmental chairman, and I didn't know what she meant—it meant nothing to me. But it does now! There is a chairman but it doesn't amount to anything but turning in some reports. But this older teacher tries to act as chairman, and the other one is much more active and gets around more and knows more people. Oh, it's quite a situation! I just try to stay neutral, that's about all I can do.

This situation resembles power conflicts seen among officers in business and government agencies. The main line of conflict is between the old and the middle-aged teacher—while the young teacher, although she may have some power in the balance, feels vulnerable and seriously in danger should she become identified with either of the other two. This situation, in its confined small-group setting, seems to be set apart from the more chronic intergenerational problem.

Problems of Recognition, Involvement, and Identity

As most mature teachers defend seniority rights, the school's traditional procedures, and traditional educational ideology, they seem to be crying out for recognition. The recognition they seek goes to the heart of the occupational role—the dedication of their full womanhood to other people's children.

I.: If you had your life to live over again, is there anything you would do differently?

48s: I don't know. Of course, I think the normal thing is for a woman to marry. I don't think there is any question about that; I think that is it. But if things don't work out that way and if things have gone differently, then the thing that is next best, if you haven't had children of your own, is to work with children. . . . But they aren't your children, either, not really . . . That's one of the things teachers have to learn. No matter how good a job you do, you can't expect much appreciation for it. Now and then, maybe, but not much. Most of them just come and go. And people think you are a funny old schoolteacher, kind of old.

Contact with students terminates with graduation in the large city, and there are difficulties involved in judging the effects of one teacher on the child's development. Thus, assessing results is a major problem for the teaching occupation. This dilemma is made severe by the totality of the unmarried teacher's occupational dedication. Like most mature teachers, 48s feels that her influence on students has been very general and intangible. Like many of her age, she reaches back into her early career for tangible evidence.

I.: What kind of influence do you think you have had on students?

48s: I don't really know. That's a funny thing to say, I suppose, after teaching all these years, but, like I say, the students come and go and you can't really tell. I suppose I have had a general influence on students, but I can't really say that I've changed their lives a great deal —or I don't know that I have.

There was one boy who I had in my class when I first came here. He was a very intelligent boy but came from a family that wasn't poor exactly but would have trouble sending a boy through school. I talked to him about going to college and helped him write to various schools and apply for scholarships. He ended up going clear through medical school and is a very successful doctor now. He wrote me for a number of years, and I have been quite proud of him.

Now, what he would have done if I hadn't talked to him about going on to college I don't know, but he is the only one I can think of that I know that I have influenced. Maybe I've had more influence on the others and I don't know about it.

A more tangible and contemporary influence on students is achieved by the "mother counselor" type of mature teacher. In the case of 56m, who devised and established a course in family living, this accomplished through a formalization of the mother-substitute function in the curriculum. In the case of 35m, the teacher who married for the first time at the age of fifty-five, it is accomplished both through extracurricular and curricular attentions to children.

> *35m:* When I say I don't have time for the children, that's not quite accurate. Hank and I have been giving special attention to one. It is the lonely children we try to look after. There was one girl who tried to commit suicide. She had lost her father and had kind of a father complex. We used to take her for rides and things like that. I think Hank was more important than I was. One time we were going to have her to dinner and I wasn't feeling well so Hank took her out to dinner. I think that meant a lot to her. I think a married teacher—and I haven't been married very long—is able to do more with children that way, if her husband is sympathetic, of course. But she doesn't need us any more now. She's doing all right. . . .
>
> There was a boy last year in my class who had been in an automobile accident and had had a broken back. This interfered with his motor activities. Between me and the principal, we got him through. And I think he earned it. We finally told him that he must turn in every paper I assigned and that I would return it to him with comments for corrections and then he would turn it in again corrected. We carried through on this. I don't think I ever felt so moved about a child as I did when he graduated. I think it was a wonderful thing that he was able to make the comeback that he did.

It is interesting that 35m, who seemed to approach marriage as an appendage to a teaching career, mobilized her (apparently willing) husband as an essential accessory in carrying out parentlike functions for children in special need of such attentions. But the most important point illustrated by 35m's story is that results are more tangible *outside* the routine of classroom teaching—where parentlike attention is given to children or where (as in the case of the boy with the broken back) there are special obstacles to be overcome. The routine of the high school teaching schedule—large classes for short periods—apparently does not provide teachers with proof that the dedication of their full womanhood to other people's children has been worthwhile. To get such proof seems to involve venturing beyond the routine and expanding the teaching role.

While such highly *feminine* "mother counselor" teachers (as 35m) expand the teaching role to include mother or maiden-aunt functions, *feministic* "old character" teachers (as 45s, who began teaching with primary career investments in the subject field) expand the teaching function to include a fatherly type of counsel about entering a work field.

I.: Did you find that, as a rule, you have closer contact with the boys than with the girls?

45s: Well, perhaps the boys are more loyal to a teacher and want to keep in contact with her longer than the girls. Still, I don't know whether that is true or not. One of my girls is in Hollywood with her husband and I hear from her all the time— they're with a TV program. The husband is one of the writers and she does make-up for the show. My girls do come back to see me but I believe that through the years the boys come back more.

You might be interested in ———. She is one of my girls. During the war she was with the USO. She had her own troupe of about three people. She was back here after she finished her duties with the USO. She came and had lunch with me and she was leaving the next week for New York. So I said to her, "Now, you are going to be a success in this business—you know that—and I know that, but your chief asset is your youth. When you get established, that's all right, the parts will come to you. When you get to New York, you go to a priest—she is Catholic—and you say to him that you are going to be an actress and you tell him, 'Now I want you to tell me what agent to go to and when that agent gives me an assignment I want to be able to come back to you and advise me which one to take.' The priest will not only be delighted and flattered but will feel a deep obligation. If you will do this I know you will cut off ten years in getting your success."

Well, that's exactly what she did. The agent gave her a couple of assignments and she took them to the priest and he said, "You take this one." It was in the chorus of a musical called *Park Avenue*.

Billy Rose saw her, and he said there were two extras that gave him the feeling the theater was not dead—one was an English girl who came over here in a play and one was the girl in the line in *Park Avenue*. He got an interview with her and he wrote up this story I have told you, about her being the smartest girl he had met in going to the priest. I didn't get any credit for it, but I got a thrill just the same.

Although an unusual teacher such as 45s is able to excite students throughout her career, there is (even in her case) some tendency to think more often of students she had when she was a young teacher. The great majority of teachers report only fragmentary evidence to indicate personal influence on a student and only rare expressions of appreciation or recognition. What evidence there is—usually taken from early career—sometimes becomes a sustaining prop of deep importance. For the great majority, the results of their efforts are difficult to identify. Older teachers perhaps are even more grateful than aging parents for appreciative recognition of their early efforts.

23s: That, I think, of course, is the unrewarding part of a teacher's life. Just like a minister—well even a minister has it better. You take a lawyer, doctor, he either wins his cases or he doesn't. You can take a doctor, either he cures his patients or they die. Or a minister can build up his church. But a teacher has absolutely no way of knowing when she has done a good job or a poor one—except when she gets something like that.

23s's mother: Like that young man who telephoned you?

23s: Yes. A student I had one time when I first started teaching came here to visit his brother. I had seen him once or twice. He's married, with children, living in Texas. He gave me his address and he knew that my mother was living with me and he said, "Now, this will be my permanent address and anytime, if you need me for any reason in the world, you telephone me and I'll be on the first plane."

Well, I had evidently built up in that young man a respect for me, or a liking for me, that has lasted all these years—enough so that he was willing to inconvenience himself to come if I need him to help out with any situation. I believe he'd do it. Because there is no reason in the world for him to say so if he didn't.

This teacher is an only child and has never married. Now, at sixty-two, with a mother of eighty-five, 23s faces the prospect of illness and death for her mother and relative isolation for herself. Although it seems unlikely that she will ever call on him, the young man from Texas provides some comfort as a substitute for a son or a grandson. More devoid of relatives than most teachers, 23s points up the kind of personal gratitude and recognition that old teachers would like to get from students.

It seems significant that 23s's mother brought up the young man from Texas. At various points in the interview, the mother intervened to suggest information that might have the effect of making 23s's life seem more worthwhile—as if to provide motherly reassurance to 23s or to allay guilt feelings she herself might have about being a burden to her daughter. And 23s does not hesitate to say that the responsibility does have the effect of restricting her activities.

23s: Well, I haven't been to the theater for quite some time because it's difficult for mother to get out, and I don't like to leave her alone. I belong to two of the teacher associations. I don't go to any of them though. I pay my dues, but, since mother and I are alone—I feel that since I can't be there, at least I can pay my dues and if the money helps out, well and good.

Judging from the sample interviewed for this study, about two thirds of single teachers assume responsibility for aging parents—usually mothers—at some point in their career. This responsibility comes most often when teachers are late middle-aged. Although this pattern is general enough so that friendship groups made up of teachers with aging mothers are formed, the prolonged confinement imposed by the care of aged, enfeebled mothers often has the effect of isolating teachers from colleagues and professional activities. Since the period of confinement may be long, since the pattern of colleagueship undergoes change, and since the teacher herself is older when she emerges from this confinement, it is difficult to become re-established, even in the teacher group.

50s: Losing mother made a big difference to me because it was the last thing I had. Oh, watching the boys—my two nephews—grow up and wondering how they were coming out of the war was interesting too. Of course, I worried a lot about mother the last year or two, and that made me feel I didn't want to live very long. She was eighty-three, and clear until the end she was just as bright as ever. I knew, and she did too, that the time wasn't far off. She lived a good life, and I was so grateful to her and just couldn't express my gratitude for all the good things that happened to her and to me even at the end. It does something to you; you can't go through an experience like that without it doing something to you.

I.: Do you feel that you want to live with anyone else again—friends, or . . .

50s: Well, that is an adjustment that I haven't worked out yet. I don't know what to do. I'm not very happy by myself. I find that living alone makes me restless and I run around to movies and things more than I should because I get lonesome. I thought I would not do anything for the first year and then see what I wanted to do. I feel that when you get as old as I am it is awfully hard to start living with someone else. I have lots of friends but they have apartments by themselves and I haven't seen so much of them during the time that mother was ill. This is just an adjustment that I haven't worked out yet.

Except for that given the warmly feminine mother counselors and such unusually effective sponsors as drama teacher 45s, the recognition old teachers get from students tends to be confined to respectful visits, on the order of servicemen home on leave "making the rounds." The indications are that such recognition from students declines as the teacher ages. This contributes to her isolation and encourages a sense that recognition is lacking. Bitterness, as we have seen, is apt to be directed toward the younger generation of students and teachers. It is also directed against the ungrateful public (with a special sense of helpless frustration). In all interviews where the subject was discussed, old teachers claimed that the public does not adequately appreciate teachers—that teachers, almost like a minority group, are singled out for adverse public stereotyping.

I.: If you gave advice to a young teacher just starting out, what would it be?

23s: Find another job.

I.: You really would?

23s: I really would. I wouldn't advise any young person to go into teaching. Now there may be a good many young people who are interested in children, who feel that they want to teach. I would say, in the first place, a teacher gets neither enough pay nor respect. You're considered sort of a prude and no one is interested in having you in a social group. You're just something apart . . . One of our teachers at school was quite disturbed about the teacher's picture they had on TV—a woman with her hair done up on top of her head and a stick in her hand. She was just supposed to be an "old schoolmarm," that's all.

I.: How do you account for this?

23s: Well, I think it's sort of a carry-over from the time when teachers were misfits. You remember the famous statement of Shaw's. "Those who *can* do, those who *can't* teach?" Of course, that's about the most untrue statement that was ever made, but that is the general attitude of people as a whole.

That attitude has been carried over and has never been gotten rid of. And it will never be gotten rid of until two things are done: until they respect a teacher as they do a doctor or a lawyer, scientist, or a business man—when they commence to consider that teaching is a profession and give teachers that respect; and when they pay them accordingly. Then they may attract the right kind of people into teaching. Until that time you're going to find that it is the unusual kind who can't do anything else, as has been said.

Included in 23s's statement are the major elements that teachers resent in the public's concept of them—that they are prudish, social misfits, "a woman with her hair done up on the top of her head and a stick in her hand." Also included in the statement is the teacher's somewhat wishful identification with the male professions, an identification that teachers and educators have sought for some time.

For many reasons, it is difficult to determine the influence of the old maid teacher stereotype on general morale, on modes of interaction with outsiders, on concepts of self, and in personal identities. The adverse stereotype may be viewed as somewhat similar to ethnic stereotypes as an instrument for labeling teachers as a group. To some extent it may operate as a self-fulfilling prophecy —even for those who react most strongly against it. Whether or not 23s, for example, would be selected as a teacher out of a crowd of women of her age is quite problematical; yet, even while protesting against the "old schoolmarm" stereotype, she appeared, to the interviewer, to fit the image. Like Jews and Negroes, teachers who fit the stereotype are noticed more than those who do not, and characteristics of teachers that reflect the stereotype are noticed more often than those which do not.

Despite some counter attitudes, it is young teachers upon whom favorable public sentiments seem to be centered. While stereotypes of Negroes, for example, are age and sex graded in such a way that aversion is centered on the young adult male, unfavorable attitudes toward teachers seem to center on the older unmarried woman. Age and sexlessness are built into the phrase "old maid school teacher," but "young teacher" is employed with implications of warmth, understanding, competence, dedication, and self-reliance. The fact that the public image of teachers should be age-polarized in this fashion may derive from an aversion to women who do not fulfill normal sex-role developmental tasks. It may also derive from students and parent reactions to young and old teachers. (It may be uncomfortable for parents to turn over their children to someone so far removed in age from the students' peer group.)

However this may be, old teachers generally are sensitive and embittered about the old maid teacher stereotype. Although 23s concedes there may be misfits in teaching (as a result of the adverse stereotype of the occupation), she is typical of older teachers in her willingness to cast the blame on the public. Young teachers were inclined to view the situation in what may be called "vicious circle" terms. They also were inclined to introduce the optimistic possibility that teachers might rally to reconstruct the public image. A thoughtful and sensitive spokesman for this perspective is 37s.

> *37s:* I have no solution. It just seems to me that it is a vicious circle. You don't have enough good teachers for them to command, in their own right, respect and admiration; and you won't get them until you build up that respect and admiration and attractiveness as far as salary and tenure and all that stuff is concerned. And I don't see how you are going to build that up until your public is satisfied that the teachers are doing what they should, and how are they going to do what they should until—I really don't have any solutions.
>
> It is worthwhile for those of us that are interested not to give them any more ammunition and not to be apologetic about our jobs. Too few teachers even want people to know they teach school. I think if they were proud of us it would make a lot of difference, so I would like it if teachers could assert that proudness in action more than in words, or else not do it.

Along with a "let's hold our heads up" drive among young unmarried teachers, there is a movement to maintain and expand extrateacher contacts. They appear to be making a special effort to fend off the imperialistic occupational identity. As thirty-three-year-old 12s observed, "When I meet people for the first time and they ask me what I do, I say I will tell them, if they won't call me a schoolteacher." Another approach—the carrying of an extra job in order to have an extrateacher diversion—is suggested by 38s.

38s: I've wanted to do so many things. Our time gets so crowded that we don't have the time to do all the things we want to do. I am interested in doing other things—away from school. Last summer I worked in a bank, a little suburban bank, and enjoyed it tremendously. I have kept my contacts there all winter and have helped out some, but financially I can hardly afford it. They don't pay enough to keep a churchmouse, but I'll go back next summer again. I work there on Friday afternoons when it doesn't interfere with my school work.

Another young teacher described, with considerable pride, her association with a high school girl in the neighborhood.

13s: I think it is good to get outside your profession. You see a different side. For instance, we have a young couple across the street, and they have a freshman daughter. She has been coming over and having me help her with her English at night. But I have learned more than she has because I get the feel of the pupil outside the classroom, and her reaction as to what goes on, which has been most interesting.

While 13s introduced this as an example of "getting outside of the profession," the teacher role is definitely involved in the relationship. Although—as mature teachers observe with annoyance—young teachers try to cast off the occupational role when the bell rings, it appears that this is not easily accomplished.

The young teacher's drive to separate the work role from other areas of her life may be a new trend of some significance. As our society places increasing emphasis on leisure and its use (and as we employ more married teachers), it would seem to follow that the teacher's occupational badge would be de-emphasized. However, the drive seems a little late for most of the teachers in this study—especially such teachers as forty-four-year-old 2s, who followed the rural route to the city and, once there, became "fixed" in a schoolteacher world.

I.: What are your plans for the future?

2s: Well, I think I'll stay here. It doesn't pay financially or any other way to move around too much. And I'm not so outstanding that I'm going to get an offer from the Department of Education at the University of Chicago or anything like that. No, I think I will stay here, and I hope to avoid getting into a rut. I hope to meet new friends and to do different things on summer vacations. That's always fun.

I.: What would you advise a young teacher just starting out?

2s: I'd advise her to be interested in a greater variety of things than I have been interested in. Mine have been confined a little too much. For example, I was never interested in anything athletic. I think it is good for people when they are young to go in for tennis or bowling or something like that. It makes a better-rounded person.

And she should try to participate more—have a wider circle of friends than I have had. I don't know how to do that. She should try to get where you can meet a greater variety of people—not just schoolteachers. I don't know how to work that out, but maybe the young people can. Oh, I've got friends I've met at church and other places but I think it is important and stimulating to have a wider circle of friends.

And then I would advise her that she is in a profession she can be proud of. And that depends on you—you can make it something to be proud of or ashamed of.

Summary

In describing their early teaching experiences, teachers gave a warm and glowing account of "having a wonderful time," of being close to students, and of securing enthusiastic response from them. Many teachers attributed their satisfaction with the job to age: "I wasn't much older than they were." Teachers seem to recognize a short orientation period, a "best teaching period," while still young and vigorous, and a decline beginning at the age of thirty-five or forty. Almost without exception, teachers who were in their thirties at the time of interviewing expressed concern about losing some kind of intimate, informal contact with students. In comparison, middle-aged teachers seemed secure, relaxed, and appeared to have accepted increased distance from students—having established for themselves a parentlike authority role. Old teachers seemed frustrated and tended to complain about students—"getting worse every year." They vacillated between two themes: "Teaching is for the young and vigorous," and "The younger generation is undisciplined, uninterested, morally decadent."

There are two types of teachers who, despite advanced age, fall outside this pattern. These are the "old characters," whose paramount interest is the subject matter and who focus on students with interest and promise in the subject field. To some degree, they seem to maintain the interest and respect of students because of extrateacher status. There are also the "mother counselors," who make use of grandmotherly understanding and who, despite their age, utilize progressive methods.

The public school system is plagued by controversy, the most basic between traditional and progressive teaching methods. Although long exposed to progressive education in summer school and through educational literature, most older teachers hold (somewhat defensively) to traditionalism. This may have to do with the mold in which they were set as young teachers. Perhaps they abandoned what progressive experimentation they may have attempted as their age distance from students increased. The young teachers in the sample, trained during a period when progressive education was being introduced, generally align themselves with the progressive school. The two ideologies often are used for name-calling purposes in a broader pattern of conflict between older and young teachers. This is an indication of the degree of teachers' involvement in the teaching role and a measure of their identification with the occupational group.

Although there were instances in which young teachers received guidance from older colleagues, the sample as a whole showed intragenerational ties to be much stronger than intergenerational ties. Most cases of conflict reported in the interviews involved an old and a young teacher. There was implication that the young teacher was fighting a cause for her entire age group, while the old teacher was upholding seniority rights and general "old guardism." In an occupation characterized by formal equality of offices, an early career ceiling, and a multitude of other status anxieties, such seemingly minor privileges as room assignments have become important prestige symbols. A critical point in the teacher's career seems to occur in the thirty-to-forty age period. This involves an intensification of interest in professional activities, a loss of intimacy with students, and a nearly complete acceptance of the fact of spinsterhood. Although it seems unlikely that such changes take place abruptly, all respondents indicated they occurred in the thirties.

This stage in the career is also likely to affect relationships within the colleague group. Among the dominant older unmarried teachers, one of the basic prerequisites for full colleagueship was spinsterhood. This suggests a total commitment to both the occupation and stereotype of "old maid teacher," a commitment that helps to generate the stereotype and builds further defensiveness among teachers themselves.

Perhaps because of older teachers' difficulties in securing students' response (as well as public aversion to women who do not fulfill the normal sex-role development tasks) unfavorable public attitudes toward teachers seem to be centered on older unmarried women. Most older teachers are sensitive to, and embittered about, the old maid schoolteacher stereotype and seem to be crying out for some form of recognition. The recognition they seek goes to the heart of the occupational role—the dedication of their full womanhood to other people's children. For the great majority of teachers there is only fragmentary evidence to indicate that they have personally influenced students and there are only rare expressions of appreciation and recognition.

Associated with this situation are pressures (lasting throughout the career) that would involve a teacher more deeply in the occupational role and that would exclude her from outside participation. These are greater in the case of teachers reared in rural areas who—in the course of coping with new problems in the urban setting—enter into friendships with teachers of their age, develop an identity wholly focused upon the occupation, and remain essentially removed from other aspects of urban experience.

Among young teachers, however, there is a drive to cast off the occupational role when the school bell rings. This attitude is in keeping with the general trend in our society, but it comes too late to affect most of the teachers in this study.

Implications for Teacher Effectiveness

These findings suggest complex relationships involving the institutional setting, career patterns, and the effectiveness of teachers. Both direct and indirect implications should be examined.

It has been suggested that age may determine relationships between teachers and students, and, therefore, the effectiveness of teachers. The findings suggest that a more complex relationship relates these two variables. Teachers' careers tend to fall into two or three phases. The first is characterized by close relationships between teachers and students and by teacher energy investment in classroom-related activities. Middle-aged teachers tend to be involved in professional activities. Older teachers often are embittered and defensive, but those who become "old characters" or "mother counselors" retain—or even increase—their ability to affect students.

Such findings imply that teachers of all ages may be effective (or ineffective), but that different types of behavior are required for effectiveness at different ages. The middle-aged teacher who yearns for the warmth of her early relations with students is probably less successful than her colleague who adjusts her behavior to her age. The elderly teacher is likely to end her career in frustrated ineffectivenes unless she becomes a subject-matter expert or a grandmother figure. Hypotheses such as these should be explored in supplementary studies, which utilize careful measurements of teacher style and student reactions.

Findings also suggest that a change is taking place in the pattern of career commitment. Young teachers are likely to be married and unwilling to commit 100 percent of their life experience to the profession. What are the effects of commitment upon effectiveness? To what extent is the new teacher a better performer than her predecessor? To what extent may conflict in the institution and the negative stereotype associated with "old maid teacher" be eliminated in order to develop a more normal kind of career commitment? Such questions are raised by the findings, not answered by them.

One practical implication should be stressed, however. Previous views of teacher role have seen the teacher in ageless abstraction. Exhaustive discussions of teacher effectiveness stress methods, sensitivity to students, and control of the classroom. Studies of teacher-pupil interaction usually differentiate among teachers in terms of subject matter, grade, or other criteria functional to school tasks. It is the contention here that such views, recommendations, and studies ignore the important facts of aging, commitment, and job satisfaction. Whichever of the variables of education one may observe, one cannot ignore the complex social worlds, the interpersonal relationships, and the career perspectives of teachers. However talented individuals may be, they will not perform to capacity in adverse social settings. The school is a *social* system; it is more than an aggregate of individuals. It is, moreover, a social system in which individuals develop, mature, age, and perform with various degrees of effectiveness at various career points and at various points of change in their social worlds.

That the effectiveness of individuals varies in terms of career points and the social settings in which they work is true of all organizations—factories, hospitals, government offices. What is special about schools is *children*. It is reasonable to expect and to insist that teachers be adequately compensated in terms of salaries and other regular increments. But, how important for teacher effectiveness are adequate rewards for the interpersonal investment in many children over many years? This is a highly vital area that remains largely unexplored.

BIBLIOGRAPHY

Ackerman, W. I. "Teacher Competence and Pupil Change." *Harvard Educational Review* 24: 273-89; 1954.

Adams, W. C. T. "Superintendent's Rating of Teachers." *Journal of Education* 90: 288-98; 1919.

Adorno, T. W., Frenkel-Brunswik, Else, Levinson, D. J., and Sanford, R. N. *The Authoritarian Personality.* New York: Harper & Row, 1950.

Alexander, W. M. *Are You a Good Teacher?* New York: Holt, Rinehart and Winston, Inc., 1959.

Allport, G. W., and Odbert, H. S. "Trait-names: a psycho-lexical study." *Psychological Monographs* 47: 1936.

———, Vernon, P. E., and Lindzey, G. *Study of Values.* (Third Edition) Boston: Houghton Mifflin, 1960.

Almy, H. C., and Sorenson, H. *Almy-Sorenson Rating Scale for Teachers.* Bloomington, Ill.: Public School Publishing Co.

Amatora, Sister M. "Highs and Lows in the Teacher Personality." *Journal of Educational Research* 48:693-98; 1955(a).

———. "Self-Appraisal in Teacher Personality." *Journal of Educational Psychology* 46: 94-100; 1955(b).

American Association of School Administrators, Department of Classroom Teachers of the National Education Association, and National School Boards Association. *Who's A Good Teacher?* Washington, D.C.: Authors, 1961.

American Management Association. "Whatever Became of Merit Rating?" *Personnel* 34(4): 8-18; 1958.

Amidon, E., and Flanders, N. A. "The Effects of Direct and Indirect

Teacher Influence on Dependent-Prone Students Learning Geometry." *Journal of Educational Psychology* 52: 286-91; 1961.

Anderson, E. W. "Techniques of Research Used in the Field of Teaching Personnel." *Review of Educational Research* 4: 15-20; 1934.

Anderson, H. H. "The Measurement of Domination and of Socially Integrative Behavior in Teachers' Contacts with Children." *Child Development* 10: 73-89; 1939.

———, and Brewer, Helen M. "Studies of Teachers' Classroom Personalities. I: Dominative and Socially Integrative Behavior of Kindergarten Teachers." *Psychological Monographs* 6; 1945.

———, and Brewer, J. E. "Studies of Teachers' Dominative and Integrated Contacts on Children's Classroom Behavior." *Psychological Monographs* 8; 1946.

———, ———, and Reed, Mary F. "Studies of Teachers' Classroom Personalities. III: Follow-Up Studies of the Effects of Dominative and Integrative Contacts on Children's Behavior." *Psychological Monographs* 11; 1946.

Anderson, H. M. "Study of Certain Criteria of Teaching Effectiveness." *Journal of Experimental Education* 23: 47-71; 1954.

Anderson, J. P. "Student Perceptions of Teacher Influence." Unpublished doctoral dissertation, University of Minnesota, 1960.

Andrews, J. H. M., and Brown, A. F. "Can Principals Exclude Their Own Personality Characteristics When They Rate Their Teachers?" *Educational Administration and Supervision* 45: 234-242; 1959.

Aschner, Mary J. "The Analysis of Classroom Discourse: A Method and Its Uses." Unpublished doctoral dissertation, University of Illinois, 1958.

Bales, R. F. *Interaction Process Analysis*. Cambridge, Mass.: Addison-Wesley, 1950.

———. "Task Roles and Social Roles in Problem-Solving Groups." In Eleanor E. Maccoby, T. M. Newcomb, and E. L. Hartley (Editors), *Readings in Social Psychology*. (Third Edition) New York: Holt, Rinehart and Winston, Inc., 1958, pp. 437-447.

———, and Strodtbeck, F. L. "Phases in Group Problem Solving." *Journal of Abnormal and Social Psychology* 46: 485-496; 1951.

Barker, R. G. "Ecology and Motivation." In *Nebraska Symposium on Motivation*. Lincoln, Neb.: University of Nebraska Press, 1960, pp. 1-50.

———, and Gump, P. V. *Big School, Small School: High School Size and Student Behavior*. Stanford, Calif.: Stanford University Press, 1964.

———, and Wright, H. F. *Midwest and Its Children*. New York: Harper & Row, 1955.

Barr, A. S. *Characteristic Difference in the Teaching Performance of Good and Poor Teachers of the Social Studies*. Bloomington, Ill.: Public School Publishing Co., 1929.

———. "Recruitment for Teacher Training and Prediction of Teaching Success." *Review of Educational Research* 10: 185-190; 268-270; 1940.

———. "The Measurement and Prediction of Teacher Efficiency: A Summary of Investigations." *Journal of Experimental Education* 16: 203-283; 1948.

———. "Teaching Competencies." In W. S. Monroe (Editor), *Encyclopedia of Educational Research*. New York: Macmillan, 1950.

———. "The Measurement of Teacher Characteristics and Prediction of Teacher Efficiency." *Review of Educational Research* 22: 169-174; 1952.

———. "The Measurement and Prediction of Teaching Efficiency." *Review of Educational Research* 10: 182-184; 1940. 13: 218-223; 1943. 16: 203-208; 1946. 19: 185-190; 1949. 25: 261-269; 1955.

———. "The Assessment of the Teacher's Personality." *School Review* 68: 400-408; 1960.

———, and Emans, L. M. "What Qualities are Prerequisite to Success in Teaching?" *Nation's Schools* 6(3): 60-64; 1930.

———, Eustice, D. E., and Noe, E. J. "The Measurement and Prediction of Teacher Efficiency." *Review of Educational Research* 25: 261-269; 1955.

———, and Jones, R. E. "The Measurement and Prediction of Teaching Efficiency." *Review of Educational Research* 28: 256-264; 1958.

———, Worcester, D. A., Abell, A., Beecher, C., Jensen, L. E., Peronto, A. L., Ringness, T. A. and Schmid J. Jr. "Wisconsin Studies of the Measurement and Prediction of Teacher Effectiveness: A Summary of Investigations." *Journal of Experimental Education* 30: 5-156; 1961.

Barrington and Company, Organization and Personnel Division. In T. G. Tulin (Editor). *Practical Approaches to Teacher Performance*

Appraisal (Merit Pay). Proceedings of a three-day seminar. New York: 1961.

Bayroff, A. G., Haggerty, H. R., and Rundquist, E. A. "Validity of Ratings as Related to Rating Techniques and Conditions." *Personnel Psychology* 7: 93-113; 1954.

Bechtoldt, H. P. "Problems in Establishing Criterion Measures." In D. P. Stuit (Editor). *Personnel Research and Test Development in the Bureau of Naval Personnel.* Princeton, N.J.: Princeton University Press, 1947, pp. 357-379.

Becker, H. S. "Role and Career Problems of Chicago Public School Teachers." Unpublished doctoral dissertation, University of Chicago, 1951.

———. "Social Class Variations in Teacher-Pupil Relationship." *Journal of Educational Sociology* 25: 451-465; 1952.

———. "The Teacher in the Authority System of the Public School." *Journal of Educational Sociology* 27: 128-141; 1953.

Beecher, D. E. *The Evaluation of Teaching: Backgrounds and Concepts.* Syracuse, N.Y.: Syracuse University Press, 1949.

———. *The Teaching Evaluation Record.* Buffalo, N.Y.: Education Publishing Co., 1953.

Bendig, A. W. "Ability and Personality Characteristics of Introductory Psychology Instructors Rated Competent and Empathetic by Their Students." *Journal of Educational Research* 48: 705-709; 1955.

Benson, A. L. "The National Teacher Examinations in 1954." *Journal of Teacher Education* 5: 244-248; 1954.

Best, J. W. "A Study of Certain Selected Factors Underlying the Choice of Teaching as a Profession." *Journal of Experimental Education* 17: 201-259; 1948.

Bestor, A. E., *Educational Wastelands.* Urbana, Ill.: University of Illinois Press, 1953.

———. "The Education Really Needed for a Changing World." *Harvard Educational Review* 27: 1-8; 1957.

Biber, Barbara. "Problems of Values and Measures in Evaluation of Teaching." *Educational Leadership* 15: 213-217; 1958.

Bicknell, J. E. *The Prediction of Effectiveness in Secondary School Teaching: A Summary Report.* Albany, N. Y.: University of the State of New York Press, 1959.

Biddle, B. J. *The Present Status of Role Theory*. Columbia, Mo.: University of Missouri Press, 1961.

———, and Simpson, Ann. *A Program for the Processing of Ordinal Data and Computation of Significance for Selecting of Central Tendency Differences*. Columbia, Mo.: University of Missouri Press, 1961.

———, Rosencranz, H. A., and Rankin, E. F., Jr. *Studies in the Role of the Public School Teacher*, Vol. 1. *Orientation, Methods and Materials*, Vol. 2. *General Characteristics of the School Teacher's Role*, Vol. 3. *Positional Differences in Teacher Role*, Vol. 4. *Role of the Teacher and Occupational Choice*, Vol. 5. *Own and Attributed Cognitions for the Teacher*, Vol. B. *Bibliographies on Role Theory and the Role of the Teacher*, Vol. D. *Bibliographies on Role Theory and the Role Methodology and Propositions*. Columbia, Mo.: University of Missouri Press, 1961.

Bloom, B. S. "Thought Processes in Lectures and Discussions." *Journal of General Education* 7: 160-169; 1953.

———. (Editor) *The Taxonomy of Educational Objectives*. New York: Longmans, 1956.

Bloomington Public Schools. *Teacher Growth Patterns*. Bloomington, Ill.: Board of Education, 1957.

Borg, W. R. "Comparison between a Performance Test and Criteria of Teacher Effectiveness." *Psychological Reports* 2: 111-116; 1956.

———. "Personality and Interest Measures as Related to Criteria of Instructor Effectiveness." *Journal of Educational Research* 50: 701-709; 1957.

Bowers, N. D., and Soar, R. S. *Studies of Human Relations in the Teaching-Learning Process. V; Final Report: Evaluation of Laboratory Human Relations Training for Classroom Teachers*. Cooperative Research Project No. 469, U.S. Office of Education, 1961.

Boyce, A. C. "Methods of Measuring Teachers' Efficiency." *14th Yearbook, National Society for the Study of Education*, Part 2. Chicago: University of Chicago Press, 1915.

Brogden, H. E., and Taylor, E. K. "The Theory and Classification of Criterion Bias." *Educational and Psychological Measurement* 10: 159-186; 1950.

Brownell, W. A. "Criteria of Learning in Educational Research."

Improving Educational Research. Official Report, Atlantic City, N.J., February 21-24, 1948. Washington, D.C.: American Educational Research Association, a department of the National Education Association, 1948, pp. 106-112.

Bureau of National Affairs, Inc. *Merit Rating of Rank-and-File Employees.* Washington, D.C.: Author, 1957.

Burnett, R. W. "The Diagnostic Problem Solving Proficiency of Elementary School Teachers in Teaching Reading." Unpublished doctoral dissertation, Indiana University, 1961.

Butsch, R., L. C. "Teacher Rating." *Review of Educational Research* 1: 99-107; 1931.

Calabria, F. M. "Characteristics of Effective Teachers." *Educational Research Bulletin* 39: 92-100; 1960.

California Teachers Association. *Teacher Competence: Its Nature and Scope.* Burlingame, Calif.: Author, 1955.

———. Research Department. *Information on Merit Rating of Teachers.* Research Bulletin 98. Burlingame, Calif.: Author, 1956.

———. Commission on Teacher Education. *Teacher Competence, Its Nature and Scope.* San Francisco: Author, 1957.

Campbell, D. T., and Stanley, J. C. "Experimental Design for Research on Teaching." In N. L. Gage (Editor). *Handbook of Research on Teaching.* Chicago: Rand McNally, in press.

Canton Public Schools. *The Canton Evaluation Plan.* Collinsville, Conn.: Board of Education, 1960.

Cassel, R. N., and Johns, W. L. "The Critical Characteristics of an Effective Teacher." *Bulletin of the National Association of Secondary-School Principals* 44: 119-124; 1960.

——— and ———. "Basic Principles of Effective Teaching Based on Adjectives Used by School Principals in Teachers' Efficiency Reports." *Peabody Journal of Education* 36: 169-173; 1961.

Castetter, D. D.; Standlee, L. S.; and Fattu, N. A. *Teacher Effectiveness: An Annotated Bibliography.* Bulletin of the Institute of Educational Research, School of Education, Vol. 1, No. 1. Bloomington, Ind.: Indiana University Press, 1954.

Central School Boards Committee for Educational Research. *Identifying Superior Teacher.* New York: Institute of Administrative Research, 1951.

Chichester, J. B. "Who Is the Good Teacher?" *Phi Delta Kappan* 37: 245-247; 1956.

Christensen, C. M. "Relationships between Pupil Achievement, Affect-Need, Teacher Warmth, and Teacher Permissiveness." *Journal of Educational Psychology* 51: 169-174; 1960.

Christenson, P. E. "Work Sampling: A Stroboscopic View of Teaching." *Educational Administration and Supervision* 42: 230-243; 1956.

Cincinnati Public Schools. *Appraisal Procedures: Appraising the Performance of Professional Personnel in the Cincinnati Public Schools.* Cincinnati, Ohio: Board of Education, 1956.

———. Department of Administration and Personnel. *Appraising the Performance of Professional Personnel in the Cincinnati Public Schools.* Cincinnati, Ohio: Author, 1959.

Coffman, W. E. "Determining Students' Concepts of Effective Teaching from Their Ratings of Instructors." *Journal of Educational Psychology* 45: 277-286; 1954.

Cogan, M. L. "Theory and Design of a Study of Teacher-Pupil Interaction." *Harvard Educational Review* 26: 315-342; 1956.

———. "The Behavior of Teachers and the Productive Behavior of Their Pupils: I. Perception Analysis; II. Trust Analysis." *Journal of Experimental Education* 27: 89-124; 1958.

Cole, D. L. "The Prediction of Teaching Performance." *Journal of Educational Research* 54: 345-348; 1961.

Conant, J. B. *Science and Common Sense.* New Haven, Conn.: Yale University Press, 1951.

Cook, W. W., Leeds, C. H., and Callis, R. *Minnesota Teaching Attitude Inventory.* New York: The Psychological Corporation, 1951.

———, and others. "Significant Factors in Teachers' Classroom Attitudes." *Journal of Teacher Education* 7: 274-279; 1956.

Cooper, J. E. "Let's Take Inventory." *Education* 80: 311-315; 1960.

Crawford, Jean M. "Impact of Activities on Participant Behavior of Children." Unpublished master's thesis, University of Michigan, 1957.

Cureton, E. F. "Validity." In E. F. Lindquist (Editor). *Educational Measurement.* Washington, D.C.: American Council on Education, 1951, pp. 695-793.

Davies, Lillian S. "Some Relationships between Attitudes, Personality Characteristics, and Verbal Behavior of Selected Teachers." Unpublished doctoral dissertation, University of Minnesota, 1961.

Davis, A. J., and Derbigny, I. A. "Toward Objectivity in Faculty Ranking." *School and Society* 81: 137-139; 1955.

Davis, C. C., and Davis, T. G. *Cooperative English Test, Test C_2 Reading Comprehension, Higher Level, Form* Z. Princeton, N.J.: Educational Testing Service, Cooperative Testing Division, 1952.

Davis, Hazel, and others. "Merit Salary Schedules for Teachers." *Journal of Teacher Education* 8: 125-197; 1957.

Della Piana, Gabriel M., and Gage, N. L. "Pupils' Values and the Validity of the Minnesota Teacher Attitude Inventory." *Journal of Educational Psychology* 46: 167-178; 1955.

Dodge, G. W., and Clifton, D. O. "Teacher-Pupil Rapport and Student Teacher Characteristics." *Journal of Educational Psychology* 47: 364-371; 1956.

Domas, S. J., and Tiedeman, D. "Teacher Competence: An Annotated Bibliography." *Journal of Experimental Education* 19: 101-218; 1950.

Donovan, F. *The Schoolma'am*. Philadelphia: Lippincott, 1938.

Downey, L. W. *The Task of Public Education: The Perceptions of People*. Studies in Educational Administration, No. 7. Chicago: Midwest Administration Center, University of Chicago, 1960.

Dyck, A. J. "Children's Social Interaction at School: A Comparison of an English and an American School and a Comparison of Teachers and Classmates." Unpublished manuscript, Midwest Psychological Field Station, University of Kansas, 1961.

———. "The Social Contacts of Some Midwest Children with Their Parents and Teachers." In R. G. Barker (Editor). *The Stream of Behavior*. New York: Appleton-Century Crofts, 1963.

Ebel, R. L. "Using Tests for Evaluation." *National Elementary Principal* 35: 29-31; 1955.

Educational Testing Service, Cooperative Test Division. *Sequential Tests of Educational Progress, Form 3A. Mathematics*. Princeton, N.J.: Author, 1957.

Ellena, W. J., Stevenson, Margaret, and Webb, H. V. (Editors). *Who's A Good Teacher?* Evanston, Ill.: American Association of School Administrators, Department of Classroom Teachers of the National Education Association, and National School Boards Association, 1961.

Ellis, J. R. "Relationships between Aspects of Preparation and Measures of Performance of Secondary Teachers of the Social Studies." *Journal of Educational Research* 55: 24-28; 1961.

Fattu, N. A. "Common Sense Versus Experimental Inference in Educational Research." *Growing Points in Educational Research.* Washington, D.C.: American Educational Research Association, a department of the National Education Association, 1949, pp. 175-181.

Fawl, C. L. "Disturbances Experienced by Children in Their Natural Habitats." Unpublished doctoral dissertation, University of Kansas, 1959.

Festinger, L. *A Theory of Cognitive Dissonance.* New York: Harper & Row, 1957.

Filson, T. N. "Factors Influencing the Level of Dependence in the Classroom." Unpublished doctoral dissertation, University of Minnesota, 1957.

Flanagan, J. C. "The Critical Incident Technique." *Psychological Bulletin* 51: 237-258; 1954.

Flanders, N. A. "Teacher-Pupil Contacts and Mental Hygiene." *Journal of Social Issues* 15(1): 30-39; 1959.

———. "Diagnosing and Utilizing Social Structures in Classroom Learning." In National Society for the Study of Education, *The Dynamics of Instructional Groups.* 59th Yearbook, Part II. Chicago: University of Chicago Press, 1960(a), pp. 187-217.

———. "Interaction Analysis in the Classroom: A Manual for Observers." Unpublished manuscript, University of Michigan, 1960(b).

———. *Teacher Influence, Pupil Attitudes and Achievement: Studies in Interactions Analysis.* Final Report, Cooperative Research Project No. 397, Minneapolis: University of Minnesota, 1960(c).

———. *Interaction Analysis: A Technique for Quantifying Teacher Influence.* Paper presented to the American Educational Research Association, February 23, 1961. Washington, D.C.: a department of the National Education Association, 1961.

Foshay, A. W. "Educational Goals." *Educational Research Bulletin* 36: 29-36; 1957.

French, W., and others. *Behavioral Goals of General Education in High School.* New York: Russell Sage Foundation, 1957.

Fry, E. B. "Research Tools: Instrumentation in Educational Research." *Review of Educational Research* 30: 513-521; 1960.

Fulkerson, G. "A Résumé of Current Teacher Personnel Research." *Journal of Educational Research* 47: 669-681; 1954.

Furse, B. S. "Merit Pay is Feasible and—Sometimes—Desirable." *Phi Delta Kappan* 42: 143-147; 1961.

Gage, N. L. Address appearing in "Proceedings," *Research Résumé*, 1960, 16. Burlingame, Calif.: California Teachers Association.

———. "The Appraisal of College Teaching: An Analysis of Ends and Means." *Journal of Higher Education* 32: 17-22; 1961.

———, and Cronbach, L. U. "Conceptual and Methodological Problems in Interpersonal Perception." *Psychological Review* 62: 411-422; 1955.

———, Leavitt, G. S., and Stone, G. C. *Teachers' Understanding of Their Pupils with Pupils' Ratings of Their Teachers.* Psychological Monograph No. 406. Washington, D.C.: American Psychological Association, 1956.

———, and Orleans, J. S. "Guiding Principles in the Study of Teacher Effectiveness." *Journal of Teacher Education* 3: 294-298; 1952.

———, Runkel, F. J., and Chatterjee, B. B. *Equilibrium Theory and Behavior Change: An Experiment in Feedback From Pupils to Teachers.* Urbana, Ill.: Bureau of Educational Research, University of Illinois, 1960.

Gallie, N. B. "What Makes a Subject Scientific?" *British Journal for the Philosophy of Science* 8: 118-140; 1957.

Gardner (Mass.) Public Schools. *A Self-Rating Scale for Teachers.* Boston: Houghton Mifflin, 1947.

Getzels, J. W., and Guba, E. G. "The Structure of Roles and Role Conflict in the Teaching Situation." *Journal of Educational Sociology* 29: 30-40; 1955.

Gibb, C. A. "Classroom Behavior of the College Teacher." *Educational and Psychological Measurement* 15: 254-263; 1955.

Gibson, R. C. "Paying for Pedagogical Power." *Phi Delta Kappan* 42: 148-151; 1961.

Gnagey, W. J. "Effects on Classmates of a Deviant Student's Power and Response to a Teacher-Exerted Control Technique." *Journal of Educational Psychology* 51: 1-8; 1960.

Goodenough, Eva E. "The Forced-Choice Technique as a Method for Discovering Effective Teacher Personality." *Journal of Educational Research* 51: 25-31; 1957.

Goodwin, J. *Principles and Practices of Teacher Evaluation Programs in the Elementary Schools of the United States.* Research Bulletin No. 97. Burlingame, Calif.: California Teachers Association, 1956.

Gowan, J. C. "Prediction of Teaching Success: Rating by Authority Figures." *California Journal of Educational Research* 6: 147-152; 1955(a).

———. "Relation of the 'K' Scale of the MMPI to the Teaching Personality." *California Journal of Educational Research* 6: 208-212; 1955(b).

———. "Use of the Adjective Check List in Screening Teaching Candidates." *Journal of Educational Research* 49: 663-672; 1956.

———. "Self-Report Tests in the Prediction of Teaching Effectiveness." *School Review* 68: 409-419; 1960.

———, and Gowan, May S. "Teacher Prognosis Scale for the MMPI." *Journal of Educational Research* 49: 1-12; 1955.

Gross, N.; Mason, W. S.; and McEachern, A. W. *Explorations in Role Analysis.* New York: Wiley, 1958.

Grotke, E. M. "A Study of Professional Distances between the Raters of Teachers and Teachers Rated." *Journal of Experimental Education* 24: 1-41; 1955.

Guba, E. G., and Getzels, J. W. "Personality and Teacher Effectiveness: A Problem in Theoretical Research." *Journal of Educational Psychology* 46: 330-344; 1955.

Guilford, J. P. "Three Faces of Intellect." *The American Psychologist* 14: 469-479; 1959.

———, and Marrifield, P. R. "The Structure of Intellect Model: Its Uses and Implications." *Reports from the Psychological Laboratory,* No. 24. Los Angeles: University of Southern California Press, 1960.

Gump, P. V., and Kounin, J. S. "Issues Raised by Ecological and "Classical" Research Efforts," *Merril-Palmer Quarterly of Behavior and Development* 6: 145-153; 1959-1960.

———, Schoggen, P., and Redl, F. "The Camp Milieu and Its Immediate Effects." *Journal of Social Issues* 13(1): 40-46; 1957.

———, and Sutton-Smith, B. "Activity-Setting and Social Interaction: A Study." *American Journal of Orthopsychiatry* 25: 755-760; 1955(a).

———, and ———. "The 'It' Role in Children's Games." *The Group,* 17: 3-8; 1955(b).

Hale, P. P. "Isolating Objective Factors for the Teaching Profession." *Journal of Educational Research* 48: 497-507; 1955.

Harlow, H. F. "The Formation of Learning Sets." *Psychological Review* 56: 51-65; 1949.

Harrison, S. M. *The Social Survey.* New York: Russell Sage Foundation, 1931.

Hart, H. L. A., and Honoré, A. M. *Causation in the Law.* Oxford: Clarendon Press, 1959.

Hartshook, E. A. "Merit Rating—Key to Better Pay and Better Teaching." *Phi Delta Kappan* 42: 157-160; 1961.

Haskew, L. D. "The Discipline of Education and America's Future." Pittsburgh, Pa.: University of Pittsburgh Press, 1959.

———. "The Teaching Profession and Professional Education." *Educational Forum* 22: 21-29; 1957.

Heil, L. M., Feifer, I., and Powell, Marion. "Characteristics of Teacher Behavior Related to Achievement of Children in Several Elementary Grades." Unpublished manuscript, Brooklyn College Book Store, 1960.

Henderson, E. "Florida's Career Increment Plan." *Phi Delta Kappan* 42: 152-153; 1961.

Howsam, R. B. *New Designs for Research in Teacher Competence.* San Francisco: California School Boards Association and California Teachers Association, 1960(a).

———. *Who's A Good Teacher: Problems and Progress in Teacher Evaluation.* Prepared for the Joint Committee on Personnel Procedures of the California School Boards Association and the California Teachers Association. Burlingame, Calif.: 1960(b).

Hughes, E. C. "Institutional Office and the Person." *American Journal of Sociology* 43: 404-413; 1937.

Hughes, Marie M., and others. "Assessment of the Quality of Teaching in Elementary Schools." Unpublished research report, University of Utah, 1959(a).

———, and others. *Development of the Means for the Assessment of the Quality of Teaching in Elementary Schools.* Unpublished research report, University of Utah, 1959(b).

Hutchins, R. M. *The Conflict in Education in a Democratic Society*. New York: Harper & Row, 1953.

Jayne, C. D. "A Study of the Relationship between Teaching Procedures and Educational Outcomes." *Journal of Experimental Education* 14: 101-134; 1945.

Jelke, F. *Efficiency*. Athens, Ohio: Ohio University Press, 1907.

Jensen, A. C. "Determining Critical Requirements for Teachers." *Journal of Experimental Education* 20: 79-85; 1951.

Jensen, O. C. *The Nature of Legal Argument*. Oxford: Blackwell, 1957.

Johnson, G. B., Jr. "An Evaluating Instrument for the Analysis of Teacher Effectiveness." *Journal of Experimental Education* 23: 331-344; 1955.

———. "An Experimental Technique for the Prediction of Teacher Effectiveness." *Journal of Educational Research* 50: 679-689; 1957.

Jones, Margaret L. "Analysis of Certain Aspects of Teaching Ability." *Journal of Experimental Education* 25: 153-180; 1956.

Kagan, J., and Moss, H. A. *Birth to Maturity: A Study in Psychological Development*. New York: Wiley, 1962.

Katz, I. "Why I Oppose Selective Merit Pay." *Phi Delta Kappan* 42: 161-163; 1961.

Kaufman, H. *The Efficient Age*. New York: Doubleday, 1913.

Kearney, N. C. *Elementary School Objectives*. New York: Russell Sage Foundation, 1953.

———, and Rocchio, P. D. "The Relation between the Minnesota Teacher-Attitude Inventory and Subject Matter Taught by Elementary Teacher." *Educational Administration and Supervision* 41: 358-360; 1955(a).

———, and ———. "Relation between a Teacher-Attitude Inventory and Pupils' Ratings of Teachers." *School Review* 63: 443-445; 1955(b).

———, and ———. "Using the Minnesota Teacher-Attitude Inventory in Counseling Prospective Teachers." *Personnel Guidance Journal* 34: 156-160; 1955(c).

Kelley, E. L., and Fiske, D. W. *The Prediction of Performance in Clinical Psychology*. Ann Arbor, Mich.: University of Michigan Press, 1951.

Kellogg, P. *The Pittsburgh District*. New York: Russell Sage Foundation, 1909.

Kerlinger, F. N. "The Mythology of Educational Research: The Methods Approach." *School and Society* 88: 149-155; 1960.

King, L. A. "The Present Status of Teacher Rating." *American School Board Journal* 70: 44-46; 1925.

Kinney, L. B. "Self-Evaluation: The Mark of a Profession." *Educational Leadership* 15: 228-231; 1958.

Knoell, Dorothy M. "A Second Attempt to Predict Teaching Success for Word Fluency Data." *Journal of Educational Research* 49: 13, 25; 1955.

Knox, W. B. "A Study of the Relationships of Certain Environmental Factors to Teaching Success." *Journal of Experimental Education* 24: 95-141; 1956.

Kounin, J. S., and Gump, P. V. "The Ripple Effect in Discipline." *Elementary School Journal* 59: 158-162, 1958.

———, ———, and Ryans, J. J. "Explorations in Classroom Management." *Journal of Teacher Education* 12: 235-246; 1961.

Kowatrakul, S. "Some Behaviors of Elementary School Children Related to Classroom Activities and Subject Areas." *Journal of Educational Psychology* 50: 121-128; 1959.

Lamke, T. A. "Prevalance and Content of Graduate Courses in the Methodology of Educational Research." *Journal of Educational Research* 49: 103-111; 1955.

Lansing Public Schools. *To Teacher Approaching a Self-Improvement Survey Year*. Lansing, Michigan: Board of Education, 1955-1956.

Levin, H. "A New Perspective on Teacher Competence Research." *Harvard Educational Review* 24: 98-105; 1954(a).

———. "New Perspectives on Teacher Competence Research." *Journal of Experimental Education* 24: 361-367; 1954(b).

———, Hilton, T. L., and Leiderman, G. F. "Studies of Teacher Behavior." *Journal of Experimental Education* 26: 81-91; 1957.

Lewin, K. *Field Theory in Social Science*. New York: Harper & Row, 1951.

Lieberman, G. F., Hilton, T. L., and Levin, H. "Studies of Teachers' Behavior: A Summary Report." *Journal of Teacher Education* 8: 433-437; 1957.

Lieberman, M. *Education as a Profession*. Englewood Cliffs, N.J.; Prentice-Hall, 1956.

———. *The Future of Public Education*. Chicago: University of Chicago Press, 1960.

Lindsey, Margaret (Editor). *New Horizons for the Teaching Profession.* Washington, D.C.: National Commission on Teacher Education and Professional Standards, National Education Association, 1961.

Long, Sister Mary. "A Synthesis of Recent Research Studies on Predicting Teaching Efficiency." *Catholic Educational Review* 55: 217-230; 1957.

Lovell, G. D., and Haner, C. F. "Forced Choice Applied to College Faculty Rating." *Educational and Psychological Measurement* 15: 291-304; 1955.

McCall, W. A. *Measurement of Teacher Merit,* Publication No. 284. Raleigh, N.C.: State Superintendent of Public Instruction, 1952.

McCulloch, R. W. "The Role of Graduate Schools in Teacher Education: A Study of Ten Graduate Programs." *North Central Association Quarterly* 30: 211-218; 1955.

MacDonald, F. J. *Educational Psychology.* Belmont, Calif.: Wadsworth, 1959.

Macfarlane, M. J., and Furso, B. S. *An Experimental Study of Teacher Evaluation.* Sandy, Utah: Jordan School District, 1959.

Mann, R. D. "A Review of the Relationships between Personality and Performance in Small Groups." *Psychological Bulletin* 56: 241-270; 1959.

Mathis, C. "The Relationship between Salary Policies and Teacher Morale." *Journal of Educational Psychology* 50: 275-280; 1959.

Mead, Margaret. *The School in American Culture.* Cambridge, Mass.: Harvard University Press, 1951.

Means, R. K. "A Teacher Appraisal Scale." *Journal of Health, Physical Education, Recreation* 31: 36-37, 1960.

Medley, D. M., and Klein, A. A. "Measuring Classroom Behavior with a Pupil-Reaction Inventory." *Elementary School Journal* 57: 315-319; 1957.

———, and Mitzel, H. E. "Pupil Growth in Reading: An Index of Effective Teaching." *Journal of Educational Psychology* 48: 227-239; 1957.

———, and ———. "A Technique for Measuring Classroom Behavior." *Journal of Educational Psychology* 49: 86-92; 1958.

———, and ———. "Some Behavioral Correlations of Teacher Effectiveness." *Journal of Educational Psychology* 50: 239-246; 1959.

———, and Williams, Ida F. *Predicting Teacher Effectiveness with the Minnesota Multiphasic Personality Inventory.* New York City, Board of Higher Education Research Series No. 34. 1957.

Megel, C. J. "Merit Rating is Unsound." *Phi Delta Kappan* 42: 154-156; 1961.

Mehl, B. "The Teacher: A Skilled or Professional Worker." *School and Society* 83: 186-187; 1956.

Millen, C. "Influence of Camp Activity-Settings on Impulsivity in the Disturbed Child." Unpublished Master's thesis, Wayne State University, 1958.

Miller, N. E. "Liberalization of Basic S-R Concepts: Extensions of Conflict Behavior, Motivation, and Social Learning." In S. Koch (Editor), *Psychology: A Study of a Science.* Vol. II. New York: McGraw-Hill, 1959.

Minneapolis Public Schools. *Goals for Performance.* Minneapolis, Minn.: Board of Education, 1959.

Mitchell, J. B. "Merit Rating: Past, Present, and Perhaps." *Phi Delta Kappan* 42: 139-142; 1961.

Mitzel, H. E. *A Behavioral Approach to the Assessment of Teacher Effectiveness.* New York: Office of Research and Evaluation, Division of Teacher Education, College of the City of New York, 1957.

———. "Teacher Effectiveness." *Encyclopedia of Educational Research.* (Third Edition) New York: Macmillan, 1960, pp. 1481-1485.

———, and Gross, C. F. *A Critical Review of the Development of Pupil Growth Criterion in Studies of Teacher Effectiveness.* Research Series No. 31. New York: Office of Research and Evaluation, Division of Teacher Education, College of the City of New York, 1956.

———, and Medley, D. M. *Studies of Teacher Behavior: Pupil Growth in Reading—An Index of Effective Teaching.* New York: Board of Higher Education Research Series No. 32. 1956.

———, and Rabinowitz, W. *Assessing Social-Emotional Climate in the Classroom by Withall's Technique.* Psychological Monographs 67, No. 18. Washington, D.C.: American Psychological Association, 1953(a).

———, and ———. *Reliability of Teachers' Verbal Behavior: A Study of Withall's Technique for Assessing Social-Emotional Climate in* *the Classroom*. Publication 15. New York: College of the City of New York, Division of Teacher Education, Office of Research and Evaluation, 1953(b).

———, ———, and Ostreicher, L. M. "The Effects of Response Sets on the Validity of the Minnesota Teacher Attitude Inventory." *Educational and Psychological Measurement* 16: 501-515; 1956.

Monroe, W. S. *Teaching-Learning Theory and Teacher Education, 1890 to 1950*. Urbana, Ill.: University of Illinois Press, 1952.

Monroe, W. S., and Clark, J. A. "Measuring Teacher Efficiency." University of Illinois Bulletin, *Educational Research Circular* No. 25. Urbana, Ill.: University of Illinois Press, 1924.

Morre, C. H., and Cole, D. "The Relation of MMPI Scores to Practice Teaching Ratings." *Journal of Educational Research* 50: 711-716; 1957.

Morsh, J. E. "The Q-Sort Technique as a Group Measure." *Educational and Psychological Measurement* 15: 390-395; 1955.

———. *Systematic Observation of Instructor Behavior*. Report No. AFPTRC-TN-56-22. San Antonio, Tex.: USAF Personnel and Training Research Center, 1956.

———, Burgess, G. G., and Smith, P. N. *Student Achievement as a Measure of Instructor Effectiveness*. San Antonio, Tex.: USAF Personnel Training Research Center, Report No. AFPTRC-TN-55-12, 1955.

———, ———, and ———. "Student Achievement as a Measure of Instructor Effectiveness." *Journal of Educational Psychology* 47: 79-88; 1956.

———, and Wilder, E. W. *Identifying the Effective Instructor: A Review of Quantitative Studies, 1900-1952*. Research Bulletin No. AFPTRC-TR-54-44. San Antonio, Tex.: USAF Personnel and Training Center, 1954.

Moss, F. A., and others. *Teaching Aptitude Test*. (Research Instrument), George Washington University, 1927.

Nagel, E. *The Structure of Science*. New York: Harcourt, 1960.

National Education Association. *Proceedings*. Washington, D.C.: Author, published annually.

———, American Educational Research Association. "Teachers and Nonacademic Personnel." *Review of Educational Research* 22: 163-272: Author, 1952.

———, National Commission on Teacher Education and Professional

Standards. *Measures of Teacher Competencies.* Report of Special Group D, the Miami Beach Conference. Washington, D.C.: Author, 1953.

———, National Commission on Teacher Education and Professional Standards. *Factors in Teaching Competence.* Report of Special Group A, the Albany Conference. Washington, D.C.: Author, 1954.

———, National Commission on Teacher Education and Professional Standards. *Defining Teacher Competencies.* Report of Special Group A, the DeKalb Conference. Washington, D.C.: Author, 1955.

———, Research Division. "Teacher Personnel Practices, Urban School Districts, 1955-56." *Special Memo,* June 1956. Washington, D.C.: Author, 1956.

———, Research Division. *Teachers' Salaries: Materials Available from the National Education Association.* Washington, D.C.: Author, 1957.

———, Research Division. "Quality of Service Provisions in Salary Schedules, 1958-59." *Research Report* 1959-R24. Washington, D.C.: Author, 1959.

———, Department of Classroom Teachers. *I Look at Me.* (Revised Edition) Washington, D.C.: Author, 1961.

———, Research Division. "Estimates of School Statistics, 1961-62." *Research Report* 1961-R22. Washington, D.C.: Author, 1961.

———, Research Division. "Why Have Merit Plans for Teachers' Salaries Been Abandoned?" *Research Report* 1961-R3. Washington, D.C.: Author, 1961.

National Industrial Conference Board. *What Employers Are Doing for Employees.* New York: Author, 1936.

Nation's Schools. "Extra Pay for Superior Teaching: Opinion Poll." Author 57: 92; 1956.

Neiman, L. J., and Hughes, J. W. "The Problem of the Concept of Role-A Re-Survey of the Literature." *Social Forces* 30: 141-149; 1951.

Nelson, K. G., and others. *Development and Refinement of Measures of Teaching Effectiveness.* Albany, N.Y.: State Education Department, 1956.

New England School Development Council. *Teacher Competence and Its Relation to Salary.* D. V. Tiedeman (Editor) Cambridge, Mass.: Author, 1956.

New York State School Boards Association, Central School Boards Committee for Educational Research. *Identifying Superior Teachers.* New York: Teachers College, Columbia University, 1959.

New York State Teachers Association. Committee on Merit Payments. *Teacher Merit and Teacher Salary.* Albany, N.Y.: Author, 1947.

———. *Progress Report on Continuing Survey of Research on Teacher Evaluation.* Albany, N.Y.: Author, 1959.

Norfolk City Public Schools. *A Guide to Teacher Growth through Self-Evaluation.* Norfolk, Va.: Board of Education, 1953.

Ohio State University, College of Education. *The Ohio Teacher Record: The Anecdotal Observation Form.* (Second Revised Edition) Columbus, Ohio: Ohio State University Press, 1945.

Ohlsen, M. M., and Schultz, R. E. "Projective Test Response Patterns for Best and Poorest Student Teachers." *Educational and Psychological Measurement* 15: 18-27; 1955.

Olander, H. T., and Kleyle, H. M. "Differences in Personal and Professional Characteristics of a Selected Group of Elementary Teachers with Contrasting Success Records." *Educational Administration and Supervision* 45: 191-198; 1959.

Orleans. J. S., Clarke, D., Ostreicher, L., and Standlee, L. "Some Preliminary Thoughts on the Criteria of Teacher Effectiveness." *Journal of Educational Research* 45: 641-648; 1952.

Osgood, C. E. "The Similarity Paradox in Human Learning, A Resolution." *Psychological Review* 56: 132-143; 1949.

Patrick, T. L. "The Importance of Evaluating the Work of the Individual Teacher." *Educational Administration and Supervision* 42: 4-9; 1956.

Peck, R. F. "Predicting Principals' Ratings of Teacher Performance from Personality Data." *Journal of Educational Psychology* 50: 70-74; 1959.

Peterson, C. H. "Seven Keys to Evaluating Teacher Competence." *American School Board Journal* 136: 34-36; 1958.

Popham, W. J., and Standlee, Lloyd S. *Out-of-School Activities and Professional Performance of Teachers.* Bulletin of the School of Education, Indiana University, 1958.

———, and Trimble, R. R. "The Minnesota Teacher Attitude Inventory as an Index of General Teaching Competence." *Educational and Psychological Measurement* 20: 509-512; 1960.

Probst, J. B. *Measuring and Rating Employee Value*. New York: Ronald, 1947.

Provo City Schools, Merit Study. *Pattern of Effective Teaching*. Provo, Utah: Author, 1961.

Provus, Malcolm. "Good Teaching Wears a Label." *Illinois Education* 48: 159; 1959.

Rabinowitz, W., and Travers, R. M. W. "Problems of Defining and Assessing Teacher Effectiveness." *Educational Theory* 3: 212-219; 1953.

Racine Public Schools. *An Inventory for Teacher Improvement*. Racine, Wis.: Board of Education.

Rankin, E. F., Biddle, B. J., Sigler, J. E., and Guthrie, H. M. "A Comparative Study of Own and Attributed Cognitions for Public School Teachers." Unpublished manuscript, University of Kansas City, 1960.

Raush, H. L., Dittman, A. T., and Taylor, T. J. "Person, Setting, and Change in Social Interaction." *Human Relations* 12: 361-378; 1959.

Rayburn, S. W. "The Genesis of Rating Sheets for Department Store Training." *Journal of Retailing* 20: 34-37; 1944.

Reavis, W. C., and Cooper, D. H. *Evaluation of Teacher Merit in City School Systems*. Supplementary Educational Monographs, Number 59. Chicago: University of Chicago, 1945. (Out-of-print; available only in libraries.)

Redl, F. "The Impact of Game Ingredients on Children's Play Behavior." In B. Schaffner (Editor), *Group Processes: Transactions of the Fourth Conference*. New York: Josiah Macy, Jr., Foundation, 1959, pp. 33-82.

Remmers, H. H., Barr, A. S., Bechdolt, B. V., Coxe, W. W., Gage, N.L., Orleans, J. S., and Ryans, D. G. "Report of the Committee on the Criteria of Teacher Effectiveness." *Review of Educational Research* 22: 238-263; 1952.

———, Barr, A. S., Bechdolt, B. V., Gage, N. L., Orleans, J. S., Pace, C. R., and Ryans D.G. "Second Report of the Commitee on Criteria of Teacher Effectiveness." *Journal of Educational Research* 46: 641-658; 1953.

Riesman, D. "Thoughts on Teachers and Schools." In Melvin Lasky (Editor) *The Anchor Review*. New York: Doubleday, 1955, pp. 27-60.

Robertson, J. D. C. "Analysis of the Views of Supervisors on the Attributes of Successful Graduate Student Teachers." *British Journal of Educational Psychology* 27: 115-126; June 1957.

Rocchio, P. D., and Kearney, N. C. "Teacher-Pupil Attitudes as Related to Nonpromotion of Secondary School Pupils." *Educational and Psychological Measurement* 16: 244-252; 1956.

———, and ———. "Using an Attitude Inventory in Selecting Teachers." *Elementary School Journal* 56: 77-78; 1955.

Rommetveit, R. *Social Norms and Roles*. Minneapolis: University of Minnesota Press, 1954.

Rose, Gale. "Toward the Evaluation of Teaching." *Educational Leadership* 15: 231-238; 1958.

Ryans, D. G. "Appraising Teacher Personnel." *Journal of Experimental Education* 16: 1-30; 1947.

———. "Comparing the Qualifications of Teachers." *Teacher Selection Papers and Reports*, No. 11. New York: National Committee on Teacher Examinations, 1947.

———. "Statistical Procedures in the Selection of Teachers." *Journal of Educational Research* 40: 695-705; 1947.

———. "The Criteria of Teaching Effectiveness." *Journal of Educational Research* 42: 690-699; 1949.

———. "Teacher Personnel Research." *California Journal of Educational Research* 4: 19-27, 73-83; 1952.

———. "Investigation of Teacher Characteristics." *Educational Record* 34: 371-396; 1953.

———. "Notes on the Ratings of Teacher Performance." *Journal of Educational Research* 47: 695-703; 1954(a).

———. "Teacher Characteristics Schedule, Form E54." (Research Instrument), University of Texas, 1954(b).

———. "Theory Development and the Study of Teacher Behavior." *Journal of Educational Psychology* 47: 462-475; 1956.

———. "Notes on the Criterion Problem in Research with Special Reference to the Study of Teacher Characteristics." *Journal of Genetic Psychology* 91: 33-61; 1957(a).

———. "Research Designs for the Empirical Validation of Tests and

Inventories." *Educational and Psychological Measurement* 17: 175-184; 1957(b).

———. "A Note on Activities of Teachers during Their Childhood and Adolescence." *California Journal of Educational Research* 9: 57-59; 1958.

———. "Some Validity Extension Data from Empirically Derived Predictors of Teacher Behavior." *Educational and Psychological Measurement* 18: 355-370; 1958.

———. "Teacher Effectiveness Research: Problems and Status." *California Journal of Educational Research* 9: 148-158, 166; 1958.

———. "Measurement and Prediction of Teacher Effectiveness." In *Proceedings of Invitational Conference on Testing Problems.* Princeton, N.J.: Educational Testing Service, 1959(a), pp. 52-67.

———. "Some Correlates of Teacher Behavior When Criterion Groups Comprised of Teachers Receiving Uniformly High and Low Observer Assessments on Three Major Patterns of Teacher-Classroom Behavior Are Employed." *Educational and Psychological Measurement* 19: 3-12; 1959(b).

———. "Teacher Characteristics in Relation to Marital Status of the Teacher." *California Journal of Educational Research* 10: 68-72; 1959(c).

———. *Characteristics of Teachers: Their Description, Comparison, and Appraisal.* Washington, D.C.: American Council on Education, 1960(a).

———. "Prediction of Teacher Effectiveness." *Encyclopedia of Educational Research.* (Third Edition) New York: Macmillan, 1960(b), pp. 1486-1490.

———. "Some Concomitants of Inventory Estimated Emotional Maturity of Experienced Teachers." *California Journal of Educational Research* 11: 106-109; 144; 1960(c).

Rugg, H. O. "Is the Rating of Human Character Practicable?" *Journal of Educational Psychology* 12: 425-438, 485-501; 1921, and 13: 30-42, 81-93; 1922.

Sandiford, P., Cameron, M. A., Conway, C. B., and Long, J. A. *Forecasting Teaching Ability.* Department of Educational Research, University of Toronto Press, 1937.

Sanford, C. W., and Trump, J. L. "Preservice Selection." In Walter S.

Monroe (Editor), *Encyclopedia of Educational Research.* New York: Macmillan, 1950, pp. 1390-1395.

Sartor, L. "How Good a Teacher Am I?" *NEA Journal* 45: 448; 1956.

Scheffler, I. *The Language of Education.* Springfield, Ill.: Charles C. Thomas, 1960.

Schultz, R. E., and Ohlsen, M. M. "Interest Patterns of Best and Poorest Student Teachers." *Journal of Educational Sociology* 29: 108-112; 1955.

Scott, W. D., and Clothier, R. C. *Personnel Management.* New York: McGraw-Hill, 1923.

Scriven, M. "The Philosophy of Science in Educational Research." *Review of Educational Research* 30: 422-429; 1960.

Sells, S. B. "Psychological Aspects of Research Design." Unpublished manuscript, Texas Christian University, 1960.

Shea, J. A. *The Predictive Value of Various Combinations of Standardized Tests and Subtests for Prognosis of Teaching Efficiency.* Educational Research Monographs, Vol. 19, No. 5. Washington, D.C.: Catholic University of America Press, 1955.

Sherif, M., and Sherif, Carolyn W. *An Outline of Social Psychology.* (Revised Edition) New York: Harper & Row, 1956.

Simpson, R. H. "Use of Teacher Self-Evaluation Tools for the Improvement of Instruction." *American Association Bulletin of Colleges for Teacher Education* 8: 1-9; 1960.

———, and Seidman, J. M. *Student Evaluation of Teaching and Learning: Illustrative Items for Teacher Self-Evaluative Instruments.* Washington, D.C.: American Association of Colleges for Teacher Education, 1962.

Skinner, B. F. "Are Theories of Learning Necessary?" *Psychological Review* 57: 193-216; 1950.

———, *Verbal Behavior.* New York: Appleton, 1957.

Smith, B. O. "A Concept of Teaching." *Teachers College Record* 61: 229-241; 1960.

———, and Ennis, R. H. (Editors). *Language and Concepts in Education.* Chicago: Rand McNally, 1961.

———, and Henderson, K. B. *The Teaching of Critical Thinking.* Unpublished manuscript, University of Illinois.

Smith, D. E. P. "Fit Teaching Methods to Personality Structure." *High School Journal* 39: 167-171; 1955.

Stapley, M. E. "A Study of Teacher Effectiveness." *Teachers College Journal* 30: 41-42; 1958.

Stein, H. L., and Hardy, J. "A Validation Study of the Minnesota Teacher Attitude Inventory in Manitoba." *Journal of Educational Research* 50: 321-338; 1957.

Stevens, S. S. "Psychology and the Science of Science." *Psychological Bulletin* 36: 221-263; 1939.

Stewart, L. H. "Study of Critical Training Requirements for Teaching Success." *Journal of Educational Research* 49: 651-661; 1956.

Stiles, L. J. *Revolution in Instruction.* The Second Charles W. Hunt Lecture, Chicago, Meetings of the American Association of Colleges for Teacher Education, February 22, 1961. Washington, D.C.: The Association, 1955.

Stogdill, R. M. *Individual Behavior and Group Achievement.* New York: Oxford University Press, 1959.

Stoller, N. "The Teacher: A New Definition." *Phi Delta Kappan* 42: 133-135; 1960.

Stoops, E., and Marks, J. R. "What about Teacher Evaluation?" *School Executive* 77: 97; 1957.

Storlie, T. R. "Selected Characteristics of Teachers Whose Verbal Behavior is Influenced by an Inservice Course in Interaction Analysis." Unpublished doctoral dissertation, University of Minnesota, 1961.

Symonds, P. M. "Characteristics of the Effective Teacher Based on Pupil Evaluations." *Journal of Experimental Education* 23: 289-310; 1955.

———, and Dudek, Stephanie. "Use of the Rorschach in the Diagnosis of Teacher Effectiveness." *Journal of Projective Techniques* 20: 227-234; 1956.

Symposium on Dimensions of Stimulus Situations Which Account for Behavior Variance. Texas Christian University: Author, 1961.

Taylor, F. W. *The Principles of Scientific Management.* New York: Harper & Row, 1911.

Taylor, H. R. "Teacher Influence on Class Achievement: A Study of the Relationship of Estimated Teaching Ability to Pupil Achievement in Reading and Arithmetic." *Genetic Psychology Monographs* 7: 81-175; 1930.

Terman, L. M., and Oden, M. H. *The Gifted Group at Mid-Life.* Stanford, Calif.: Stanford University Press, 1959.

Thelen, H. A. (Editor). "Experimental Research toward a Theory of Instruction." *Journal of Educational Research* 45: 89-136; 1951.

Thorndike, R. L. *Personnel Selection.* New York: Wiley, 1949.

Thurstone, L. L., and Thurstone, Thelma G. *American Council on Education, Psychological Examination for College Freshman.* Princeton, N.J.: Educational Testing Service, Cooperative Test Division, 1952.

Tiedeman, D. V., and Cogan, M. "New Horizons in Educational Research." *Phi Delta Kappan* 39: 286-291; 1958.

Tolle, E. R., and Murray, W. I. "Forced Choice: An Improvement in Teacher Rating." *Journal of Educational Research* 51: 679-685; 1958.

Tomlinson, Loren B. "Pioneer Studies in the Evaluation of Teaching." *Educational Research Bulletin* 34: 63-71; 1955(a).

———, "Recent Studies in the Evaluation of Teaching." *Educational Research Bulletin* 34: 172-186; 1955(b).

———. "An Attempt to Develop an Instrument for Measuring Professional Understandings in Elementary Education." *Journal of Educational Research* 51: 659-667; 1958.

Tompkins, E. E., and Roe, V. "The Case For and Against Merit Rating." *Bulletin of the National Association of Secondary-School Principals* 40: 5-21; 1956.

Torgerson, T. L. "The Measurement and Prediction of Teaching Ability." *Review of Educational Research* 4: 261-266; 1934; 7; 242-246; 1937.

Toulmin, S. E. *The Uses of Argument.* Cambridge, Mass.: Cambridge University Press, 1958.

Travers, R. M. W. *An Introduction to Educational Research.* New York: Macmillan, 1958.

Turner, R. L. *Problem Solving Proficiency among Elementary School Teachers. II. Teachers of Arithmetic, Grades 3-6.* Bloomington, Ind.: Institute of Educational Research, Indiana University Press, 1960.

———. *Problem Solving Proficiency among Elementary School*

Teachers. IV. Further Investigations of Teachers of Arithmetic, Grades 3-6. Bloomington, Ind.: Institute of Educational Research, Indiana University Press, 1961.

———, and Fattu, N. A. *Problem Solving Proficiency among Elementary School Teachers. I. The Development of Criteria.* Monograph of the Institute of Educational Research. Bloomington, Ind.: Indiana University Press, 1960(a).

———, and ———. "Skill in Teaching: A Reappraisal of the Concepts and Strategies in Teacher Effectiveness Research." *Bulletin of the School of Education, Indiana University Press,* 36: 1-40; 1960(b).

Valentine, E. "Occupational Expectations of Three Normal School Students Groups." Unpublished Master's thesis, University of Chicago, 1950.

Vander Werf, L. S. "The Evaluation of Teaching." *American School Board Journal* 133: 27-30; 1956.

———. *How to Evaluate Teachers and Teaching.* New York: Holt, Rinehart and Winston, Inc., 1958.

Wade, E. *Problem Solving Proficiency among Elementary School Teachers. III. Teachers of Reading, Grades 2-5.* Bloomington, Ind.: Institute of Educational Research, Indiana University Press, 1961.

Wagenschein, Miriam. "Reality Shock: A Study of Beginning Elementary School Teachers." Unpublished Master's thesis, University of Chicago, 1950.

Waimon, M. D. "Observing the Classroom Acting System." *American Educational Research Association Newsletter* 12: 20; 1961.

Walker, Helen M. (Editor). *The Measurement of Teaching Efficiency.* New York: Macmillan, 1935.

Waller, W. *The Sociology of Teaching.* New York: Wiley, 1935.

Warner, W. L., Havighurst, R. J., and Loeb, M. B. *Who Shall Be Educated?* New York: Harper & Row, 1944.

Watters, W. A. "Annotated Bibliography of Publications Related to Teacher Evaluation." *Journal of Experimental Education* 22: 351-367; 1954.

———. "The Teacher-Appraisal Conference." *National Elementary Principal* 35: 30-44; 1956.

Webb, W. B., and Bowers, N. D. *Student Performance as a Measure of Instructional Proficiency.* Research Report Project No.

NM001077-01-06. Washington, D.C.: U.S. Naval School of Aviation Medicine, 1957(a).

———, and ———. "The Utilization of Student Learning as a Criterion of Instructor Effectiveness." *Journal of Educational Research* 51: 17-23; 1957(b).

———, and Nolan, C. Y. "Student, Supervisor, and Self-Ratings of Instructional Proficiency." *Journal of Educational Psychology* 46: 42-46; 1955.

Weber School District. *Progress Report of the Weber School District Merit Study, June 1961.* Ogden, Utah: Author, 1961.

West Hartford Public Schools. *West Hartford Teacher Appraisal Plan.* West Hartford, Conn.: Author, 1961.

Whitney, F. L. "The Social and Economic Background of Teachers College and University Students." *Education* 47: 449-456; 1927.

Winget, J. "Ecological and Socio-Cultural Factors in Teacher Inter-School Mobility." Unpublished doctoral dissertation, University of Chicago, 1951.

Wispe, L. G. "Evaluating Section Teaching Methods in the Introductory Course." *Journal of Educational Research* 45: 161-186; 1951.

Withall, J. "The Development of a Technique for the Measurement of Social-Emotional Climate in Classrooms." *Journal of Experimental Education* 17: 347-361, 1949.

———. "Assessment of the Social-Emotional Climates Experienced by a Group of Seventh Graders as They Moved from Class to Class." *Educational and Psychological Measurement* 12: 440-452; 1952.

———. "Research Tools: Observing and Recording Behavior." *Review of Educational Research* 30: 496-512; 1960.

Wright, H. F., Barker, R. G., Nall, J., and Schoggen, P. "Toward a Psychological Ecology of the Classroom." *Journal of Educational Research* 45: 187-200; 1951.

Wynn, D. R., and DeRemer, R. W. "Staff Utilization, Development, and Evaluation: Evaluation of Staff Effectiveness." *Review of Educational Research* 31: 400-401; 1961.

Yourglich, Anita. "Study on Correlations between College Teachers' and Students' Concepts of 'Ideal-Student' and 'Ideal-Teacher.'" *Journal of Educational Research* 49: 59-64; 1955.

Index

INDEX

ACE-raw score, 117–119
Achievement, 8, 202, 206–207, 213, 215
  concept of, 204
  measures of, 199
Age and generational differences
  in teaching, 283, 289, 292, 294, 299–300, 313–314
  educational ideology and, 282–283, 300
  in extracurricular activities, 288–291
  intergenerational ties, 296, 312
  intragenerational ties, 296, 312
  occupational identity and, 291–293
  in professional associations, 289
  as progressive-traditional issue, 283–284, 287
  and seniority privileges, 283, 300
  See also Spinsterhood, 287, 292–293, 312–313
Age-grading of teachers, by students, 272–276
Allport-Vernon-Lindzey Study of Values, 118
Antecedents, 234
Approach-avoidance conflicts, 110
Approach to educational experimentation, 39
A prori classification, 29
Attitude inventory, 199

Backgrounds, 247–248
Behavior, studied at different levels, 128
Behavior setting, 173–174, 183, 185–186
Behavior-setting management, 180–181
Bill of rights, of teachers, 230

Categories, episodic, 140–146
Characteristic, 244
Children's activities, 168–170

Children's behavior in school, 175–176
Criteria for classifying entries, 148–151
Classroom discourse, 136–137
Classroom environments, as behavior settings, 185–186
Classroom saliency, student perceptions of, 182
Classroom, sequence of events in, 203
Classroom situation, 17, 184
  manipulation of, 33–34
Cognitions, levels of, 246
  types of, 244–245
Community size, 270
Comprehensive interviews, 264–265
Comprehensive personnel program, 64
Consequences, long-term, 6, 14–15
  of contributions teacher makes to school and community, 14
  of differences in lives of pupils, 14
Contextual variables, 6, 15–17
Coop C_2 test of reading comprehension, 117
Criterion variable, 19
Critical incident technique, 48, 71

Degree of independence index, 185–188
Dependence, 202, 213
Developmental training, 291
Deviant, 166–168
Diagnostic Reading Problems B, 114, 116–117, 123–125
Dominative contacts, 204–205

"Early plateau" occupation, 268
Educational context, 37
Efficiency, 44–48
Efficiency society, 44
Employee ratings, 45, 48
English-social studies, combined classes, 211
"Entry," 139–140
Episode, 132–134, 140
  classification of, 139–140
Episodic discourse, 133–134

Equal-appearing intervals, method of finding, 48
Evaluation of teacher competence, 41–43, 64–66, 76
  forms for, 41–50, 58–59, 61–62
  methods for, 58–61
Explanation, types of, 152–153

Feminine "mother counselor" teachers, 304
Feministic "old character" teachers, 304
Filmstrips, 200
Formative experiences, 8–9, 31
Form D, Arithmetic, 111–115, 117–119, 121–122, 124–125
Form E, Arithmetic, 112–115, 117–119, 121–124
Form F, Arithmetic, 112–115, 117, 119

"Glossary" describing behaviors to be assessed, 72, 74
Goals of education, 3, 238–243

I/D ratio, 206, 208, 211–212, 219
Immediate effects, 6, 12–13
  and pupil behavior, 12
  and pupil properties, 13
In-service training of teachers, 200
Instruments for objective assessment, 19
Integrative contacts with the pupil, 204–205
Interaction analysis, 34–35, 198, 206–207
  categories, 197
  reliability of, 207–208
  training in, 207
Involvement, 301–303

Laboratory problems, 104
Language, 157
Leadership research, 38
Learning formats, 190–192
Logic, 157
  and language relations, 157–158
  and psychology relations, 159–162
Logical behavior, nature of, 131

Man-to-man ratings, 50
Mathematics classes, 209–215
Mechanical-explaining entries, 147–148
Merit rating, 45–46, 64, 76, 99
Middle-class occupation, four phases of, 266–267
Minnesota Teacher Attitude Inventory, 118
Monologue discourse, 33, 132, 135

Nonteachers' bill of rights, 230–231

Objective instruments, 23–25
Observation-and-assessment record form, 72
Observation techniques, 21–23

Person distinctions, 246–247
Probst scale, 46
Projects, research, 37
Psychological ecology, 33, 165
  hypothetical constructs in, 9
Public concept of teacher, 308
Pupil, attitudes, 205–206
  behavior, 31, 203–205, 216–218

Rating, 41–42, 48, 54–55, 63–64, 228
  forms, 25–27, 41, 50–54, 58
  methods, 52–53
Reading Problems W, 113–118, 122, 125
"Reality shock," 268
Recognition, evidence of, 301–302
Records, existing, 28–29
Reliabilities, 74–75, 114, 133, 137–138
Research projects, 37
  strategy, 102–103
Ripple effects, 165–166, 168, 183
Role approach, contributions of, 260–261
Role cognition, comparisons of, 248–249
Role theory, 234, 242–243

Scientific management, 43–47
Schools, variables, 8, 31
Score cards, 50
Self-reports, 27–28
Social–class mobility, 268
Social milieu, of the school system, 267
Social studies, 209, 211, 213–215
Sociology of occupations and professions, 266
Sociological frame of reference, 266
Sources of variation in response, 107
Specimen record, 22, 167, 170, 175, 178
Step Level 3, Mathematics Test, 117–118
Strategy for effectiveness, 32
Supervision, 42, 95–97
Supervisors, 42–43, 47
Supportive techniques, 166
Surveys, social and school, 44–45
Synomorphy, 174

TCS characteristic B, 77–79, 81, 83, 87
  I, 78–79, 83
  R, 77–79, 83
  R_1, 77–79, 81, 83
  S, 78–79, 83–84
  V, 78–79
  Z, 77–79, 81, 83, 87
TCS pattern X, 76–77, 79, 81, 83, 85–86
  Y, 76–77, 79, 81, 83, 85–86
  Z, 76–77, 79, 81, 83, 85–86
Teacher(s)
  age-graded cliques of, 294
  age, sex, and generations among, 36–39, 266, 294, 314–315
  assignment, 97–99
  occupational mobility of, 268
  role, 219, 264
  selection, 94–95
  sponsorship, 294

Teacher(s) (*cont.*)
  students as young teachers, 269–270
  traits, 228
Teacher behavior, 6, 121, 163–164, 234–236
  ambiguity of, 164
  expectancies of, 110–111
  clarity of, 164
  classification, 12
  cooperation, 221
  ideal expectations of, 205
  inconsistency of, 164
  in the classroom, 69
  preciseness of, 164
  problem-solving activity and, 12
  spontaneous, 219
  and teacher-pupil interaction, 10
  vagueness of, 164
Teacher characteristic, 31, 77–87
  early experience, 80
  emotional stability, 84
  of men and women, 82
  pupil behavior in classroom, 90
  relation to age, 81
Teacher characteristic schedule, 79–80, 84, 87–88, 125
Teacher characteristic study, 31–32, 68–79, 81, 84, 86, 90–93, 97
Teacher competence, 2, 10, 18, 232–234, 238–243
  defined, 18, 241
  development of, 241–242
  effects of educational process on, 237–238
  explanation of poor research about, 232
  formative factors in, 237–238
  measurement of, 241–243
  poor definitions of, 238–240
Teacher effectiveness, 20, 162–164, 180–181
  complexity of the problem of, 38
  confusions about, 3
  defined, 20
  environment for learning of, 165

Teacher influence, 205–207
  flexibility of, 204, 212, 219
  patterns used to achieve, 206
Teacher patterns, 76–86
Teacher properties, 6, 9–10
Teacher-pupil interaction, 37, 236–237
Teacher ratings, 47–48, 54
  field study methods of, 251–255
Teacher role, selected findings on, 256–258, 259
  broad stereotypes of, 256–257
  in community, 257
  and differences between respondent positions, 257
  study of, 35–36
Teachers, direct, 209, 213, 215, 218–219
  bill of rights, 230
  discipline problems of, 210
  extended questions of, 210
  indirect, 209–211, 213–219
  mature types of, 278
  statements by students about, 209
Teachers' relationships in career changes, 269
  statements by teachers concerning, 202–203
Teaching
  as problem-solving activity, 32, 104–107
  logic of, 33, 127–128
  poor, 227
  as series of overt acts, 196–197
  superior, 227
Teaching evaluation, 62
  to classify teachers, 226–229
  to improve instruction, 224–225
  problems in, 220
Teaching methods, 312
  progressive, 312
  traditional, 312
Teaching situation, 235–236
Teaching tasks, 103, 107–108
  empirical relevance of, 114

Teaching tasks (*cont.*)
  measures of other behaviors, 117
  relationships, 125
Teaching-task performance variables, 118–124
Teaching techniques, 166

Valuative dimension, 155–156
Verbal behavior in teaching, 129–130

S.C.E.R.T., West Bengal
Library
Calcutta
25/3, B. C. Road

370'78
BID